University of Alberta

Oral Proficiency Testing
for Grade Six Students
in a German-English Bilingual School
in Edmonton, Alberta

by

Daniela Gatto ©

A thesis to the Faculty of Graduate Studies and Research in
partial fulfillment of the requirements for the degree of
Master of Arts
in
Applied Linguistics
Department of Modern Languages and Cultural Studies

Edmonton, Alberta
Fall 2000

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59718-0

Canada

University of Alberta

Library Release Form

Name of Author: Daniela Gatto

Title of Thesis: Oral Proficiency Testing for Grade Six Students in a German-English Bilingual School in Edmonton, Alberta

Degree: Master of Arts

Year this Degree Granted: 2000

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

D. Gatto

10908-115 Street
Edmonton, Alberta
T5H 3L4

July 4/2000

ABSTRACT

Oral proficiency is often considered the most crucial in testing and evaluating the acquisition of a foreign language within a communicative teaching framework. Popular oral proficiency tests such as the Student Oral Proficiency Assessment (SOPA), are not entirely appropriate for all contexts, such as the German-English bilingual program in Alberta, where German is offered from kindergarten through Grade 12. Teachers expressed the need for an oral proficiency test, in particular for children in Grade Six.

Based on the results of a preliminary study an oral proficiency test was adapted from available tests for this group of students. The teacher field tested the first version of this test with three Grade Six students, and the final test was modified in view of the trial. Ultimately, this test will support the continuation of a successful German-English bilingual program. The test contains short questions, a picture description task, and a rating scale. This test is easy to administer, in addition to being a practical and reliable tool.

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled 'Oral Proficiency Testing for Grade Six Students in a German-English Bilingual School in Edmonton, Alberta' submitted by Daniela Gatto in partial fulfillment of the requirements for the degree of Master of Arts in Applied Linguistics.

Ann Penningroth

Dr. Ann Penningroth

Dr. Howard Grabois

Leila Ranta

Dr. Leila Ranta

June 30, 2000

ACKNOWLEDGEMENTS

Many people have helped before and during the production of the paper presented here. I want to thank each one of them because I certainly would not have been able to do this alone. The most important persons I want to thank are the talented children of the bilingual elementary school in Edmonton. Do not let anybody tell you you can not learn more than one language at a time. You prove them wrong and do it so well. A special thank you goes to Marion Fritz, the Grade Six teacher at the school who suggested the project, took time to meet and discuss goals, requirements, and possible test designs. Thank you also to the following: Susanne Eddie; Dr. Ed Friss, the past principal of the school; Mrs. Elvira Loewen, the present principal and the Edmonton Public School District. Another thank you goes to the committee members Dr. Howard Grabois and Dr. Leila Ranta, the latter for her editing and insights. Last but not least a big thank you to Dr. Ann Penningroth, my supervisor and editor, who was always there to help and encourage me.

TABLE OF CONTENTS

Chapter I. Introduction: Goals of the Study 1

Chapter II. Concepts and Terminology 5

2.0. Second Language (L2) Acquisition and Foreign Language (FL) Learning 5

2.1. Bilingual and Immersion Education 6

2.2. Language Testing and Assessment 10

2.3. Communicative Competence and Proficiency 15

2.4. Proficiency Testing 16

Chapter III. Oral Testing Methods 20

3.0. Test developed by ACTFL 20

3.1. Tests developed by CAL 23

3.1.1. SOPA 23

3.1.2. OPI 26

3.1.3. SOPI 29

3.1.4. TOPT 30

3.2. Tests developed by the American College Board and the ETS 32

3.3. Test developed for French Testing in Canada 36

Chapter IV. Test Development and Design 39

4.0. Preliminary Study 39

4.0.1. Methodology 39

4.0.2. Results of the Preliminary Study 41

4.0.3. Summary, Discussion, and Implications for Field Testing 48

4.1. Field Testing Preparation and First Test Design 51

4.1.1. Test Preparation 51

4.1.2. Test Design, Description, and Procedure 54

4.1.3. First Version of the Test 55

4.2. Field Testing 66

4.2.1. Methodology 66

4.2.2. Evaluation Results and Observations 67

4.3. Modifications to the First Test Version 72

Chapter V. Summary, Discussion, and Implications 76

5.0. Summary 76

5.1. Discussion 80

5.2. Implications 83

5.2.1. Implications for Bilingual Schools 83

5.2.2. Implications for Future FL-Tests 85

References 87

Appendix A The ACTFL Proficiency Guidelines – Speaking 93

Appendix B Nine-level Rating Scale for Student Oral Proficiency Assessment (SOPA). Adapted from the COPE 99

Appendix C TOPT Speaking Tasks 102

Appendix D ETS Scoring Criteria 106

Appendix E MOCAP Evaluation Sheet and Global Descriptors of Levels of Performance 108
Appendix F Preliminary Study: Instructions and Questions 110
Appendix G Rating and Letter Grading Scale for German-English Bilingual Grade Six Students. Adapted from the SOPA 112
Appendix H Field Testing including all Four Sections 115
Appendix I Grammar Checklist 122
Appendix J Additional Evaluation Criteria 123
Appendix K Final Test 124

LIST OF TABLES

Table 1 Analyses of language production 42

Table 2 Letter grading scale adapted to the SOPA 55

Table 3 Evaluation results and grades 68

CHAPTER I.

INTRODUCTION: GOALS OF THE STUDY

This thesis is a description of the development of an oral proficiency test appropriate for the Grade Six students of a German-English bilingual school in Alberta. Here, the process that led to the field testing of the first test version is described, as well as the final version of the oral proficiency test, its usefulness to teachers, parents, students, and the bilingual program.

The bilingual program in Alberta offers a program in which a non-dominant language, namely German, is used during some part of the school day. To be more specific, certain subjects, such as Health, Social Sciences ('*Heimatkunde*'), and Art are taught in both English and German (most topics are taught entirely in German, some in English), while German Language Arts is taught completely in German. In addition, students are encouraged to speak as much German as possible, for example when addressing the teacher outside of the regular instruction periods. In Canada, this type of early second language program is referred to as bilingual education in which elementary children learn through the medium of a second language. The children are assessed and evaluated in both languages in all subject areas. Moreover, their proficiency in German is tested as part of the evaluation.

The children enrolled in the German-English bilingual program have different language backgrounds. The majority are third-generation German children born in Canada, and so are one or both parents. In most cases, at least one of these children's grandparents emigrated to Canada from a German-speaking country, for the most part from Germany, but also from Austria and Switzerland. As a rule, most second- or third-generation Germans never knew how to speak German, or in many cases, have limited German language skills. However, a frequent contact with relatives in Germany is characteristic for a lot of these families.

Some of the families who send their children to this German-English bilingual program are for example Bosnian refugees who fled to Germany where their children attended school, and later immigrated into Canada. These children are typically fluent and native-like speakers of German, but by and large do not speak German at home. German is their second, and English their third language. These children are enrolled in the bilingual program with the goal to continue multilingual skills. A number of children in the bilingual program speak German at home. Even though the number of German-born immigrants to Canada has decreased, a small number of native German children is present in every grade level. However, the percentage of native German children rarely exceeds 10 percent (one to three students) in each class.

In the school year 1999 – 2000 a total of 189 students were enrolled in the elementary school reported on in this thesis. The students are divided into seven grades and eight classes: Kindergarten (K), K/Grade1 combined, Grades1, 2, and 3, Grades 3/4 combined, Grades 5, and 6. Dividing the students into eight classes instead of seven aims to keep the class sizes smaller and averages between 22 to 24 students per class. The predicted enrolment for the school year 2000 – 2001 at the same school is roughly the same.

I developed this oral test in response to a Grade Six teacher in one of the two bilingual elementary schools in Edmonton who expressed the need for an oral proficiency test, because no consistent oral test is in place for German. Such a test is necessary to help evaluate the students in all modalities involved in language learning, to be precise their reading, listening, speaking, and writing skills in German. An oral proficiency test would allow teachers to evaluate the students' oral abilities and to present students and parents with a holistic evaluation of the children's abilities in German. The goal is to design an oral proficiency test for these Grade Six students in a bilingual program that can be immediately used by the teacher, and that supports student assessment in German. In addition, this oral proficiency test can later be modified accordingly for kindergarten to Grade Five. In order to design an oral proficiency test, a series of steps need to be taken:

1. Determine the requirements necessary to design an evaluation method such as an oral proficiency test;
2. Identify oral proficiency tests presently in use;
3. Determine the appropriateness of available tests for the target group in the study reported here, Grade Six students in a German-English bilingual program;
4. Adapt existing test(s) to the target context;
5. Develop an oral proficiency test and field test it.

The concepts and terminology relevant to this study and to designing a test are presented in Chapter 2, and a variety of existing tests is surveyed and discussed in Chapter 3. As part of the discussion of these oral tests for children and adult learners of foreign languages, I will point out specific advantages and disadvantages of these tests that are relevant to the target group in this thesis. A detailed description and discussion of a preliminary study as well as field testing the first test version follow in Chapter 4. The final version of the oral proficiency test designed here is described and discussed in the last chapter, followed by a brief discussion of possible implications the test has on the bilingual language program and future foreign language tests.

CHAPTER II.

CONCEPTS AND TERMINOLOGY

Before discussing the characteristics of some or the oral test methods presently in use, and the oral proficiency test designed for the Grade Six students of this German-English bilingual school, I will explain the concepts and terminology relevant to this thesis.

2.0. SECOND LANGUAGE (L2) ACQUISITION AND FOREIGN LANGUAGE (FL) LEARNING

Within the field of second language acquisition research, there are two terms that denote the learning of a language other than a first language (L1), sometimes referred to as the 'mother tongue'. These are the terms second language (L2) and foreign language (FL). The difference between the two is the following: L2 refers to the acquisition of a second language that a speaker learns in addition to his or her L1 in the target language culture. This means, that an L2 language learner learns a second language within the 'natural' language environment (namely the language of the community, or country), as, for example, a Spanish L1-speaker in Western Canada learning English as L2. L2 with respect to English is primarily used in the North American research and teaching context.

In contrast to L2, the term FL is used within a cultural context and refers to the acquisition of a foreign language that a speaker learns in addition to his or her L1. However, as opposed to L2, the learner learns this FL outside of the target language environment, usually within his or her L1 language environment (cf. Apeltauer, 1993). Therefore, FL learning includes all language programs in a school, university, college, or other institute, outside of the target language culture, i.e., German at a Canadian university, or in a Canadian school.

Despite this distinction between the terms L2 and FL, within the North American research and teaching context the term L2 is used as the general term for non-primary language learning, even if the term FL would be more appropriate. Even though the children in the bilingual program reported here are FL learners, I will use FL and L2 interchangeably.

2.1. BILINGUAL AND IMMERSION EDUCATION

The teaching of a language other than English to North American children has a long tradition in the United States and Canada. Malakoff and Hakuta (1990) show that the U.S. has a long history of teaching foreign languages within the public schools, starting in the 1880s (p. 28). This was due to the large numbers of immigrants from many countries, mostly from Europe (ibid.). At that point in time, French was taught in French-speaking communities in Louisiana and Northern England, Spanish in New Mexico and California, and German in the Midwest and the East. Thompson et al. (1990) point out that "[a]s early as 1840, English-German bilingual schools were organized in a number of [American] cities, including Baltimore, Cincinnati, Cleveland, Indianapolis, Milwaukee, and St. Louis" (p. 23). However, during World War I, the German program was discontinued in those schools (ibid.). In the mid 1900s, German was also taught in Ohio, and other languages were taught as well, for example, Norwegian, Lithuanian, Czech, or Dutch. In the 1960s, after a decline in teaching languages other than English, during the war and in the post-war era, bilingual education gained in popularity. As a result of this demand for foreign language programs, the U.S. Congress passed the Bilingual Education Act (BEA) in 1968 (cf. Malakoff & Hakuta, 1990, p. 30 ff.).

Canada has a variety of bilingual language programs within its public as well as its Catholic school districts, with a dominant French immersion program across the country. The French immersion program was originally created in Quebec by Anglophone parents who wanted to maximize their children's job perspectives by developing bilingual competence (see Genesee, 1987, p. 4). The Canadian situation concerning bilingual education was further reinforced by the 1969 Official Languages Act, in which both English and French were declared official languages nationwide. Furthermore, just like the U.S., Canada has provided a home for diverse groups of immigrants, and many ethnic groups have stressed the idea of preserving 'old-country' cultures, such as observing traditions involving ethnic foods, music, or clothing, and preserving the use of the mother tongue. One example of such 'heritage language' preservation is the great interest and support for teaching children the language of their ancestors. Alberta was the first province in Canada "to legalize languages other than English or French as mediums of instruction" (Cummins, 1984, p. 88) in 1971, and has since offered a variety of languages in school: Cree, French, as well as bilingual programs in Arabic, German, French, Hebrew, Chinese Mandarin and Ukrainian (cf. Alberta Education, 1996, p. 24). As long as the provincial authorities consider the demand for a particular heritage language sufficient, and the program "promotes the continued vitality of ethnic cultures" (cultural maintenance) and enriches "children's educational experience" (Cummins, 1984, p. 88), Canadian provinces such as Alberta and Manitoba continue to make provisions for heritage language teaching within their public schools.

Immersion programs offered in schools are a particular form of bilingual education, in which "the second language and the first language are used to teach regular school subjects" (Genesee, 1987, p. 1), for example science. In order for a program to be regarded as immersion, at least 50 percent of the instructions must be taught in the second language. Consequently, other bilingual language programs differ from the immersion programs in the time spent speaking and using the foreign language. For example, students in a French core program "study the language for approximately 30 – 45 minutes every day," whereas students in the French immersion program "spend at least 50% of their time learning the second language through content subjects such as science, mathematics and social studies" (Dicks & Rehorick, 1995, p. 136). The goals of the programs are also different. The main goal of immersion programs is for the students to become 'functionally proficient' in the foreign language (see Curtain & Pesola, 1994). Students are functionally proficient in a foreign language if they "are able to communicate in the second language on topics appropriate to their age level" (p. 31). Total immersion programs use the foreign language for all subject areas in the early grades, and encourage, or at times even expect, their students to use the foreign language outside the classroom, e.g., during recess. According to Curtain and Pesola, the three goals of any immersion program are a) functional proficiency in the foreign language, b) "to master subject content taught in the foreign language" (p. 30), and c) to "acquire an understanding of and appreciation for other cultures" (ibid.). In bilingual programs, on the other hand, the time spent on the foreign language rarely exceeds 50 percent, and students are expected to acquire proficiency in listening, speaking, reading, and writing in the foreign language (ibid.). The expected levels of proficiency are not defined by Curtain and Pesola.

There are many different forms of bilingual and immersion programs, and differences within one program, like the differences between total immersion versus partial immersion, or different types of bilingual programs with variations in the amount of time spent on the second language. One characteristic of the bilingual classroom, as stated by Ellis (1985) is, that the bilingual classroom has a "[m]ixed focus – sometimes on form, sometimes on meaning" and that this type of language teaching set-up has a "[p]otentially strong resemblance to natural setting" (p. 151).

Lindholm and Fairchild (1990) describe a so-called 'bilingual immersion education' program, a foreign language school program in the United States. It "combines the most significant features of bilingual education (for language minority students) and immersion education (for language majority students)" (p. 126). The bilingual immersion program is constructed the following way:

> (a) […] the non-English language is used for a significant portion of the students' instructional day; (b) the program involves periods of instruction during which only one language is used; (c) both native English speakers and nonnative English speakers […] are participants and (d) the students are integrated for most content instruction. (Lindholm, 1990, p. 96)

Even though the program of the school described in my thesis is not considered a bilingual immersion program (the term used in the U.S.), it features some of those elements: (a) the target language (in this case German) is used for a significant portion of the day, and (b) there are times at which only one of the languages is used. Also, as described in Chapter 1, every class has several native German-speaking children who speak German at home and are learning English as a second language. Many of the Canadian-born students find themselves in the unusual situation of communicating with these German native speakers. This offers an opportunity to use German with peers, which is absent from French immersion programs in Canada.

According to Curtain and Pesola (1994), the students in a bilingual program should be proficient in the four modalities of listening, speaking, reading, and writing, and teachers must have the appropriate tools to evaluate the students in all four skills. However, depending on the goals of a school and its program design, evaluation tools will differ. The oral proficiency test I propose here aims to add an appropriate evaluation tool for the speaking component in German for Grade Six students in a bilingual elementary school.

2.2. LANGUAGE TESTING AND ASSESSMENT

Spolsky (1978) discusses some general problems he observed with foreign language testing in general, particularly controversies surrounding tests based on past research. According to Spolsky, "[i]n many countries, the contrast and even conflict between traditional subjective examinations and the newer objective standardized tests is still a central issue for professional and public concern and debate" (p. 216), and the issue of how to test a particular area, rather than what, is usually of more interest (see p. 217).

Even more challenging is "the problem of judging oral proficiency" (p. 224). Perron (1968) has remarked that "[t]he most difficult problems arise when trying to construct tests of ability to speak a language [...] although the ideal of a test based on free conversation is very attractive, the problems of sampling and reliable scoring are almost insoluble" (p. 115).

Two types of tests, achievement tests and proficiency tests, have been used to evaluate students' abilities in foreign language learning. According to Savignon (1985), achievement tests of the 1960s were designed to test "the instructional content of a particular course," whereas proficiency tests were "based on a theory of the abilities required to use language for communication" (p. 129). Bachman (1990) describes the differences between proficiency tests and achievement tests according to the test's relationship between a particular language theory and its course syllabus:

> We can refer to theory-based tests as *proficiency tests*, while syllabus-based tests are generally referred to as *achievement tests*. [...] [A] language proficiency test based on a theory of grammatical competence is likely to be quite similar to an achievement test based on a grammar-based syllabus, but quite different from an achievement test based on a notional-functional syllabus. (p. 71)

Therefore, the test offers evidence for being a proficiency test or an achievement test only in connection with the theory it represents. One type of achievement test is typically, but not necessarily, the testing of a grammatical form newly introduced, practiced, and drilled in class. This might be tested using a task where students are required to change active sentences into passive sentences. This allows the instructor to see if the student can manipulate a particular structure. One problem with this specific type of test is that the test is not used in a communicative context and the function of any particular grammatical form is not made salient for the learner. For example, German has specially marked grammatical forms for the subjunctive: the function of the subjunctive is to express contrary to fact conditions, unreal conditions, or wishes. However, if the context is not provided to the language learner, the learner may not understand when the form is used and consequently, the learner will most likely not use the second language appropriately in the required context. These types of tests allow the student to prepare for this test without any language use in context. Depending on the test results, the instructor may decide to repeat part of the instruction, or to further practise grammatical forms.

In designing a language test which assesses a speaker 'appropriately,' several factors must be considered. Appropriate assessment takes into consideration form, function, and use of the foreign language such as pronunciation, grammar, vocabulary, situational knowledge, cultural knowledge, language comprehension, and language production. The goal of a proficiency test is to assess all of the speaker's abilities in the foreign language, and to evaluate the speaker accordingly. A more detailed discussion on oral proficiency testing will follow in section 2.4.

In designing a proficiency test (as well as any other language test), particular criteria must be followed. Among other researchers, Bachman and Palmer (1996), and Finocchiaro and Sako (1983) agree that test validity and test reliability are the two most important factors in any test design.

Finocchiaro and Sako (1983) describe the four criteria of validity, reliability, comprehensibility, and practicality necessary for the design of any foreign language test. According to the authors, "[t]est validity is the most critical factor to be judged in [...] foreign language testing" (p. 24) and "[a] test is valid when it measures effectively what it is intended to measure" (ibid.). Furthermore, any test is only valid if the test results can be compared "with some outside or independent criterion" (ibid.), and four types of validity can be measured:

- content validity (assuring that all test items correspond to the course instructions);
- concurrent validity (assuring that the test scores correspond to an independent criterion or standard);
- predictive validity (assuring that the test results correspond with a student's previous performance);
- construct validity (assuring that the test is constructed according to its goal).

The second and third test criteria mentioned by Finocchiaro and Sako are reliability (or stability) and comprehensiveness. A test is considered reliable or stable if it produces "consistent results and give[s] consistent information" (p. 28) regardless of how many times the test is administered to a student. A test is considered comprehensive "when it contains an appropriate proportion of items from all aspects of the material to be tested" (p. 29). The last criterion Finocchiaro and Sako describe as necessary for useful test design is practicality, which contains three factors: economy (low in production and utilization cost), scorability (ease of scoring), and administrability (easy to administer).

Bachman and Palmer (1996) name reliability, construct validity, authenticity, interactiveness, impact, and practicality as basic principles for test development in general. Reliability, construct validity, and practicality correspond to the terms discussed by Finocchiaro and Sako (1983), however, authenticity, interactiveness, and impact are introduced by Bachman and Palmer (1996) as follows: authenticity refers to the need for the language used in the test to "correspond to language use in specific domains other than the language test itself" (p. 23). Interactiveness refers to the necessary degree of interactiveness between the test taker's individual characteristics and his or her "language ability (language knowledge plus metacognitive strategies [...], topical knowledge, and affective schemata in accomplishing a test task" (p. 39). This type of interactiveness of different types of knowledge a person has reflects the holistic idea of language, namely, that learning a language requires more than learning words, meaning, syntax, and grammar, but also metalinguistic strategies. Impact refers to "the various ways in which test use affects society, an education system, and the individuals within these" (ibid.). Bachman and Palmer declare that a test is only useful when it considers and meets all the above criteria.

This thesis is not a validation study, and consequently, one of the most important aspects necessary to prove a test's validity, its concurrent validity, will not be investigated. The objective of the oral proficiency test described here is not to measure its validity, reliability, comprehensibility, and practicality. The goal is to create an oral proficiency test that can immediately be used in the Grade Six of a particular bilingual elementary school, and that facilitates student assessment in German. For that purpose, the test's reliability and practicality are important factors for the design of the test, which will be described in more detail in chapters 4 and 5.

2.3. COMMUNICATIVE COMPETENCE AND PROFICIENCY

The terms communicative competence and proficiency are used in different ways in linguistic literature, and to make things worse, they are often used interchangeably. Ellis (1985) defines proficiency as follows:

> Proficiency consists of the learner's knowledge of the target language; it can be considered synonymous with 'competence'. 'Proficiency' can be viewed as linguistic competence or communicative competence. L2 proficiency is usually measured in relation to native speaker proficiency. (p. 302)

Unfortunately, this definition does not distinguish between proficiency and communicative competence, and according to Ellis, they are one and the same. Even though the speaker's communicative competence and proficiency are related, they are not simply interchangeable. Discussions on communicative competence appear in a variety of literature concerned with a second language learner's proficiency. Communicative competence underlies and is the cause for a speaker's language performance, and is by and large tested by language instructors. Communicative competence is operationalized in second language theory and language testing.

Based on Canale and Swain's (1980) distinction between grammatical competence, sociolinguistic competence, discourse competence, and strategic competence, Bachman and Palmer (1984) propose a model of communicative competence which distinguishes between a second language speaker's intuitive knowledge of a language, and his or her actual performance when using the language. Savignon (1985) also emphasizes the importance of "language in its context" (p. 131) and stresses the importance of grammar, as well as the presence of grammatical competence as a necessary element within communicative competence. She further stresses that communicative competence "includes much more than sentence-level grammar, or *grammatical competence*" (p. 130), and that other components are included, such as "sociolinguistic competence, [...] discourse competence, [...], and strategic competence" (ibid.), just as Canale and Swain (1980) five years before her. Savignon (1985) points out further that it has been difficult in the past to control for all of these factors, and that tests were not designed properly, resulting in vague and insufficient measures.

For the purpose of this thesis, no further deliberation on communicative competence and its definition will take place. Instead, I will use the term proficiency as it is used by the American Council on the Teaching of Foreign Languages, ACTFL (see section 3.0.), in the context of oral proficiency testing. In general, the second language learner's proficiency level ideally reflects his or her knowledge of the language in question. ACTFL defines proficiency as the language learners' functional speaking abilities, to be more specific, "their ability to accomplish linguistic tasks representing a variety of levels" (American Council on the Teaching of Foreign Languages, 1999). Testing a speaker's proficiency will be discussed in the following section.

2.4. PROFICIENCY TESTING

Researchers such as Byrnes (1987), Finocchiaro and Sako (1983), and Heilenman and Kaplan (1985) discuss proficiency testing extensively. Finocchiaro and Sako (1983) state that "[l]anguage proficiency tests are designed to measure control of language or cultural items and communication skills already present at the time of testing" (p. 21). Furthermore, "[p]roficiency tests measure […] language skills and competence, including structure, phonology, vocabulary, integrated communication skills, and cultural insights" (p. 22). Again, both content and context of language use are considered to be important factors. The authors stress that proficiency tests should not simply test the achieved level of formal learning, but instead, "proficiency tests serve principally to obtain measures of the degree of knowledge of the foreign language at a particular time" (ibid.).

Heilenman and Kaplan (1985) also describe and discuss the goals of proficiency testing criteria, adding that proficiency consists of three parts: function, topic or context, and accuracy (p. 60), which will be discussed below in more detail. Heilenman and Kaplan add that many proficiency tests in the past "tend to focus on learner deficiencies" (p. 57) and do not focus on the learners' abilities and the actual language learning progress. Furthermore, proficiency tests will demonstrate individual differences among students, which should always be taken into account (pp. 60-61).

I believe Byrnes (1987) offers the most precise explanation of the role of proficiency testing in second and foreign language teaching. She states that proficiency testing incorporates three distinctive perspectives (already mentioned above by Heilenman & Kaplan, 1985), namely "1) the communicative tasks or *functions* being performed; 2) the *context* both linguistic and nonlinguistic, […] and its specific *content*, and 3) the appropriateness or *accuracy* […] to create meaning" (Byrnes, 1987, p. 44). It becomes clear in this definition that "grammatical correctness is only one component in accuracy which, in turn, is only one of the three major perspectives" (p. 45). As a result, assessing a learner's level of proficiency offers a holistic representation of the learner's language learning stage at the time of testing. This includes information about the learner's oral proficiency.

As mentioned above, Byrnes (1987), Finocchiaro and Sako (1983), as well as Heilenman and Kaplan (1985) name accuracy as an essential criterion when testing a speaker's foreign language proficiency. In the past, grammatical and syntactic accuracy were the main features tested and evaluated in a speaker, mainly because grammatical and syntactic accuracy are easy to test. More recent proficiency testing has progressed to testing a speaker's communicative accuracy. Nevertheless, grammatical accuracy is believed to play a significant role in the interlocutor's ability to comprehend. Accuracy in a foreign or second language started to be investigated in the 1980s in predominantly sociolinguistic research regarding native speakers' attitudes towards grammatical errors produced by non-native speakers (cf. Galloway, 1980, Politzer, 1978). Some attitudinal studies revealed, that native speakers' attitudes towards the non-native speakers were greatly influenced by the amount of grammatical errors the non-native speakers produced. Based on these results, grammatical accuracy continues to be a major concern for language learners and instructors.

Another criterion often discussed and always strived towards in foreign language learning refers to the speaker's fluency in the foreign language. Different sources define fluency in a variety of ways. Fillmore (1979), for instance, claims that four types of fluency can be distinguished:

1.) the ability to speak continuously;

2.) the ability to speak reasonably well about a wide range of topics;

3.) the ability to speak eloquently, thoughtfully, and perfectly formed;

4.) the ability to use language creatively.

According to Fillmore, all four types of fluency depend entirely on the speaker's knowledge of language and control of language. The more knowledge and control of the foreign language, the more fluent the speaker. A speaker's fluency is therefore more difficult to measure than accuracy.

In this chapter, the terminology necessary to talk about second language education and the design of an oral proficiency test were described and included the distinction between L2 and FL. Various bilingual education programs in the U. S. and Canada were presented, and the requirements necessary for a language test to be considered valid, reliable, comprehensive, and practical were taken into account. As a final point, proficiency and oral proficiency testing were presented as they are used in this thesis.

CHAPTER III.

ORAL TESTING METHODS

Typically, language tests are based on the different learners and different goals of each language program. As discussed above (section 2.2.), each language test must meet specific criteria to become useful, as well as valid, reliable, practical, and comprehensive. The test developed for this thesis is based on existing oral proficiency tests, and in the following section I look at oral tests presently in use, and discuss their advantages and disadvantages primarily in respect to their practicality for the target group of the test presented in this thesis.

3.0. TESTS DEVELOPED BY ACTFL

The American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines are widely used in the United States. The ACTFL proficiency guidelines are based on years of experience with oral testing in governmental institutions, like the Foreign Service Institute (FSI). The FSI tested the second language abilities of its personnel, and most of the early testing was based on specific subject-knowledge. The later ACTFL guidelines (see Appendix A: The ACTFL Proficiency Guidelines – Speaking) were "based on a theory of the abilities required to use language for communication" (Savignon, 1985, p. 129) and were developed from "numerous research projects, scholarly articles, and debates" (American Council on the Teaching of Foreign

Languages, 1999), as well as practical experience in different second language classrooms.

Later ACTFL Oral Proficiency workshops concerned with drafts of the test still failed to address the issue of communicative competence (cf. Loughrin-Sacco, 1990, Sieloff, 1985), and the test has been scrutinized for many years because of this shortcoming (see Barnwell, 1993). However, it has been the only standardized oral proficiency test used in the United States for many years which enables a great variety of languages other than English to be evaluated. The earlier ACTFL criteria associate the different scoring levels with different levels of general education, e.g., the Advanced level "with affective and everyday uses of language, centered around the personal and the concrete" (Barnwell, 1993, p. 200), whereas "[t]he higher level (Superior) is linked to academic learning and intellectual discourse" (p. 201). Clearly, the wording in these evaluations reflects that each level, and consequently the test itself, is designed for adult L2 learners.

In 1998 ACTFL published proficiency guidelines for children, from kindergarten to Grade 12, designed to help teachers evaluate their younger language learners better by taking into consideration the cognitive stages in their foreign language learning. In order to evaluate children learning a second language, the rating guidelines are divided into six categories:

- comprehensibility (how well are you understood?)
- comprehension (how well do you understand?)
- language systems (how accurate is your language?)
- vocabulary (how extensive and appropriate is your language?)

- cultural awareness (how is your cultural knowledge reflected in your communication?)
- communication strategies (how do you maintain communication?)

(Swender & Duncan, 1998)

Within each of these categories three types of subdivisions (modes) are distinguished, namely an interpersonal, an interpretive, and a presentational mode. The interpersonal mode centres on the way the speakers negotiate meaning among themselves. The interpretative mode focuses on the way speakers interpret cultural meaning appropriately, and the presentational mode refers to how successful second language speakers are in facilitating listeners to understand and read "between the lines" (Swender & Duncan, 1998, p. 482). For the purpose of the test presented in this thesis, the three modes within their six categories are too specific and extensive to discuss any further. Nevertheless, these categories supply an interesting starting point for the test design presented in this thesis. These proficiency guidelines for children appear to be the first extensive categorization method that include all aspects of communicative competence, because forms (e.g., grammatical structures) and function (language in context) are evaluated equally. Until the creation of these guidelines, the major concentration was on linguistic forms.

3.1. TESTS DEVELOPED BY CAL

The Center for Applied Linguistics (CAL) was founded in the late 1950s and describes itself on its current web page as follows:

> CAL is a private, non-profit organization: a group of scholars and educators who use the findings of linguistics and related sciences in identifying and addressing language-related problems. CAL carries out a wide range of activities including research, teacher education, analysis and dissemination of information, design and development of instructional materials, technical assistance, conference planning, program evaluation, and policy analysis. (CAL, 1999)

CAL has been actively involved in a variety of language programs, including "English as a second language (ESL), immigrant education, foreign language education, language proficiency assessment, bilingual and vernacular language education, refugee education and services, language policy and planning, and cross-cultural communication" (CAL, 1999). In conjunction with ACTFL, CAL has also designed the CLEAR Oral Proficiency Exam (COPE) and the Student Oral Proficiency Assessment (SOPA) described in the following section, as well as the (Simulated) Oral Proficiency Interviews (OPI and SOPI), which will be discussed in section 3.1.2. and 3.1.3.

3.1.1. SOPA

In the 1987 – 1988 school year, CAL designed in collaboration with the Center for Language Education and Research, CLEAR, the Oral Proficiency Exam (COPE) for elementary school children in immersion and FLES (Foreign Language in the Elementary

School) programs (Gutstein & Goodwin, 1987, Rhodes & Thompson, 1990). Based on the increasing "number of language immersion programs around the world in the last two decades" and the consequent "increased interest in finding better ways to evaluate the language proficiency of young students" (Rhodes, 1996, p. 1), in 1991 CAL designed the Student Oral Proficiency Assessment (SOPA) which is based on the COPE, but the target group has been expanded. An earlier form of the SOPA was "used for oral language assessment of six-to nine-year-olds in a variety of immersion programs, including partial and total immersion" and "the instrument was adapted for use in non-immersion French, German, Japanese, and Spanish elementary school programs" (p. 2). The most recent SOPA draft (Thompson, Boyson, & Rhodes, 2000) aims "to determine students' oral proficiency and listening comprehension in a foreign language" (p. 1). However, the assessment procedure is applicable to "a variety of types of immersion programs, including partial and total immersion" (ibid.). During the SOPA, the students and the examiners are either audio- or video-recorded and the immersion version consists of four parts: "listening comprehension, informal questions, science and language usage, and story telling" (p. 2). Two examiners administer the test to two students at a time, and the test takes approximately ten to fifteen minutes to complete. Rhodes (1996) explains that "[t]he goal of the assessment is to show what the students *can* do with language, not what they cannot do" (p. 2). The examiners collect as much data as possible from the students in the interview, which allows them to ground their ratings on a large database in which patterns are easily visible. The testing is done entirely in the target language, and the SOPA has two rating scales to chose from which are, based on the ACTFL guidelines and the COPE scales. The six-level scale is designed for students in a Foreign Language in the Elementary School (FLES) program, and a nine-level scale for students in immersion programs (see Appendix B: Nine-level Rating Scale for Student Oral Proficiency Assessment (SOPA)). The six-level scale tests students' comprehension and fluency, whereas the nine-level scale tests the students' listening comprehension, fluency, vocabulary, and grammar. Rhodes & Thompson (1990) briefly describe these four categories, which were adapted for the SOPA:

> *Comprehension* refers to the ability to understand the spoken language in a range of situations, including formal, instructional situations in which there are few contextual cues to meaning as well as informal conversational situations. Comprehension also refers to being able to understand normal speech – speech that has not been adjusted in pace or repeated.
> *Fluency* refers to the rhythm and pacing of the speech produced by the student. It involves the degree to which speech is produced smoothly without hesitations and without pauses to search for vocabulary and expressions.
> *Vocabulary* refers to the student's knowledge of the words and expressions needed to communicate. This knowledge includes both the range of vocabulary used appropriately and the use of idiomatic words and phrases.
> *Grammar* refers to the accuracy of the speech used by the students in terms of word formation and sentence structure. When judging the student's accuracy, all of the following are considered: the frequency of grammatical errors, the degree to which they interfere with a listener's ability to understand, and the range or grammatical structures used by the student. (p. 82)

Although grammar is one of the four categories "the student's ability to communicate is the overall concern of the SOPA, grammar only is an issue if it interferes with that ability" (p. 9).

The SOPA was revised, tested, and validated in an extensive study at Iowa State University (Thompson, Kenyon, & Rhodes, 2000) in order "to investigate whether the instrument measures what it is supposed to measure" (p. 3), namely proficiency. The

SOPA was used with children in partial and total immersion programs, and content-based FLES programs. The six-level rating scale was used for the children in the FLES program and the nine-level rating scale for the children in the immersion programs which resulted in two separate validation studies, one for each language program. The study used the Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM; Padilla, 1994), as well as the CAL Student Self Assessment (SSA; Thompson, 1996) to validate the SOPA. The test results offered a "moderate to strong support for the validity of the SOPA's claim to assess listening comprehension and speaking ability in a second language for young learners" (Thompson, Kenyon, & Rhodes, 2000, p. 46) which in turn provide support for the effectiveness of the SOPA and, consequently, supports the application of the SOPA evaluation criteria with the Grade Six children in the study reported here.

The comprehensive nine-level rating scale of the SOPA (provided in full in Appendix B), is a valuable source for the oral proficiency rating scale design in the study presented here. In addition, the time span necessary to administer the test, ten to fifteen minutes per student, correlates with the time anticipated per student in the Grade Six of the bilingual school reported in this thesis .

3.1.2. OPI

The Oral Proficiency Interview (OPI) is an oral test that was first developed by ACTFL. In its *Tester Training Manual* (Buck, Byrnes, & Thompson, 1989, pp. 1-1 to 2-15), the OPI is defined as "a standardized procedure for the global assessment of functional speaking ability, or oral proficiency" (p. 1-1), and it aims to determine the speaker's patterns of strengths and weaknesses, as well as a "level of consistent functional ability" (ibid.). The OPI is not an achievement test, but "assesses language performance in terms of the ability to use the language effectively and appropriately in real-life situations" (ibid.). The OPI was designed for second language learners who learn a foreign language in a classroom setting outside of the natural language environment. The test consists of a 10-30 minute tape-recorded conversation between the tester and the testee, which should resemble a natural conversation as much as possible. The following four specific criteria are measured and assessed:

- the global tasks or functions performed with the language;
- the social context and the content areas in which the language can be used;
- the accuracy features which define how well the speaker performs the tasks pertinent to those contexts and content areas, and
- the oral text types – from individual words to extended discourse – produced. (ibid.)

The tester can adjust and personalize the questions during the conversation according to the speaker's "individual background, life experiences, interests, and opinions" (p. 1-2). The OPI is well-designed, but unfortunately, it is not designed for everyone: certified testers must conduct it, and these testers require special training in order to learn how to use the test and how to evaluate speakers accordingly.

The scoring sheet for the OPI divides speakers into four levels, following the ACTFL rating scale (for details, see Appendix A). The four levels are recapped in ACTFL's 'Inverted Pyramid':

> SUPERIOR: Can support opinion, hypothesize, discuss abstract topics, and handle a linguistically unfamiliar situation.
> ADVANCED: Can narrate and describe in past, present and future time/aspect, and handle a complicated situation or transaction.
> INTERMEDIATE: Can create with language, ask and answer simple questions on familiar topics, and handle a simple situation or transaction.
> NOVICE: No functional ability; speech limited to memorized material. (p. 2-4)

Loughrin-Sacco (1990) criticizes the OPI and the ACTFL OPI workshops extensively. He states that the "OPI is not altogether a proficiency test because test-taking tips could significantly enhance performance" (p. 708) and moreover, "the tests do not adequately measure overall communicative ability" (p. 710). Most importantly, "[c]ommunication involves more than" (p. 712) what is tested in the OPI. Despite the test's shortcomings, Loughrin-Sacco judges the OPI to be "the best comprehensive test" (p. 710) he has seen that evaluates a foreign language learner's oral skills.

The OPI is not compatible with one key issue in the case of the Grade Six students in this thesis, namely the amount of time the teacher can spend with each student. Clearly, 30 minutes per student are not viable for schoolteachers to spend on oral examination alone. With approximately 25 students in the class, there is no time for this type of task for elementary teachers in addition to evaluating all other school subjects. In addition, Canadian teachers do not have the financial resources to travel to the United States to be specially trained to apply the test. Yet, elementary teachers are highly trained professionals, and it should be possible to design an oral proficiency test in a way that any Grade Six teacher can follow the instructions and the scoring key.

3.1.3. SOPI

CAL also designed the Simulated Oral Proficiency Interview (SOPI), which attempts to mimic the OPI, in order to test large groups of language learners efficiently and reliably without a trained tester. Kuo and Jiang (1997) analyze the OPI and the SOPI and compare their advantages and disadvantages. The SOPI is, in contrast to the OPI, administered to the language learner via a tape-recorder, and does not require a specifically-trained tester to be present to administer the test. It takes approximately 45 minutes to complete the test in the full version, and since it is a taped test, the SOPI content does not vary according to the individual learner. However, for the purpose of the test for the Grade Six students the bilingual school in Alberta, this inflexibility is not desired, since students come from different language exposure-backgrounds. Kuo and Jiang state that the SOPI with its standardized design "is more suitable in cases where there are numerous interviewees who possess similar language abilities" (p. 511). Due to the lack of topic and question flexibility, the SOPI is not applicable for the students of the bilingual school in question. Furthermore, the technology (multiple recording facility) necessary to administer the SOPI to a large number of students is financially unfeasible for the bilingual elementary school.

SOPI topics are very restricted and since the teachers of the bilingual school in Alberta deal with different children and different interests, the choice of topic could have a differential effect on performance due to varying levels of prior knowledge. All Grade Six students should have an equal opportunity to talk about a topic in which they are well-versed so that they can perform as well as possible. For example, if the topic happens to be of no interest to a student, or even more so, if a student does not have any knowledge of the topic presented, he or she will not be able to talk about the topic. Consequently, the data the teacher receives will be insufficient, and will most likely misrepresent the student's oral proficiency. The teacher must elicit enough data to evaluate the students' abilities in the foreign or second language, and the more engaging the topic is for a student, the more likely it is that the student will produce a substantial amount of data, including a variety of structures, and vocabulary. It is mainly for this last reason that the SOPI is inadequate for the Grade Six students here.

3.1.4. TOPT

Stansfield and Kenyon (1991) present the final report for the Texas Oral Proficiency Test (TOPT), a testing program which is based on the SOPI and designed to certify "French, Spanish, and bilingual education teachers" (ibid.: abstract). The test contains four parts:

1.) warm-up: testee is asked several personal background questions;

2.) picture-based items (5 pictures): testee must

    - give directions
    - describe familiar setting, e.g., school, home
    - describe set of routine events
    - describe amusing event from the past
    - describe event planned for the future;

3.) topic items: testee must speak about a variety of topics, e.g., how to use public transportation;

4.) situation items: testee must respond to real-life situations, e.g., give advice or lodge a complaint.

According to Stansfield and Kenyon all of the instructions must be clear, the pictures must be appropriate, and the topics and situations must be clear and appropriate in order to allow the "examinees to demonstrate their skills" (p. 36). Even though the TOPT is designed for teachers, many of the 38 possible speaking tasks listed in the TOPT report are applicable to children at the Grade Six level, e.g., "7. Talk About Personal Activities: Be able to talk about your leisure activities, favorite pastimes, and preferred hobbies" (Appendix A, p. 6). The wording is intended for adults and necessary changes have to be made for children. The question could be rephrased to: "Tell me what you do in your free time, what your favourite pastimes are, and what kind of hobbies you have." The majority of the 38 tasks (see Appendix C, TOPT Speaking Tasks) can be modified to be appropriate for children.

Even though the target group for this test is teachers, the test design itself, and the four parts which comprise the test are of interest to the oral proficiency test design for the Grade Six students in the study reported here. The four parts described above are similar to the task used in the preliminary study that was conducted for this study, and will be described in detail in section 4.0. (Preliminary Study).

3.2. TESTS DEVELOPED BY THE AMERICAN COLLEGE BOARD AND THE ETS

In the United States, the College Board (1993; created in 1955) and the Educational Testing Service (ETS) design placement tests for foreign languages, including French, German, and Spanish, which have been used by several universities and colleges in the United States since the 1980s. "*The 1992 Advanced Placement Examination in German Language*" (College Board, 1993) offers a detailed description of the test, as well as its grading and scoring standards. Even though the test is tailored to "students' transition from secondary school to college" (p. 1), certain parts of the scoring guides are applicable to a much younger student target group, such as the Grade Six students in my study.

The German ETS test consists of two sections (I and II), and the examination totals approximately 180 minutes. Section I (multiple-choice) consists of four parts:

- Part A: Listening Comprehension - 40 questions in German (recorded materials, written responses);
- Part B: Vocabulary - 15 questions (written responses);
- Part C: Grammar and Syntax - 20 questions (written responses);
- Part D: Reading Comprehension - 25 questions (written responses).

  (p. 3)

Section II consists of free-responses, and is divided into two parts:

- Part A: Writing - cloze paragraph (fill in the blanks) and 1 essay question (written responses);

- Part B: Speaking - 6 directed responses and 1 picture stimulus (taped responses). (ibid.)

A committee consisting of five teachers from secondary schools, colleges, and universities is responsible for designing the parts of the exam. These teachers are appointed by the ETS and ETS members are also involved in reviews and revisions. The first section, the multiple choice questions, are easy to score (computer scanning), and the first part of section II (paragraph completion) is easy and fast to correct by any teacher checking an answer key. However, in order to correct and grade the essay question of Part A in section II, the student's "mastery of the language by a show of creativity" (p. 5) is taken into consideration, but no grading guidelines are supplied by ETS for this particular part. The instructions given to the students (in English) state that the composition (written in German) must be "well-organized and coherent" and that the students' "work will be evaluated for grammatical accuracy, variety, and range, as well as appropriateness of vocabulary and idioms, spelling, and organization" (p. 32).

Section I of the examination takes approximately 160 minutes to complete and consists strictly of testing the learner's comprehension, reading, and especially writing skills. Therefore, only Part B of section II, the speaking part, is of interest for the study reported here. The instructions for the speaking part are administered by tape (see comprehension part in section I) and the students' responses are tape-recorded. The speaking part consists of two subsections, one requiring six short responses and the other a picture description. The directions on the tape inform the student that he or she "will hear a number of instructions [...] or questions" that must be answered and to "address the question fully and directly" (p. 44). The student's "ability to express [...] [him- or her]self fluently and correctly will be the main basis" (ibid.) to grade the responses. The questions are stated twice each time in German, and some of the example questions are: "Nummer 1 … Beschreiben Sie eine Person, die Sie gut kennen!" ('Number 1 ... Describe a person that you know well!'), "Nummer 3 … Was haben Sie gestern alles gemacht?" ('Number 3 ... What did you do yesterday?'), "Nummer 4 ... Du bist ja heute so fröhlich! Was ist denn Schönes passiert?" ('Number 4 ... You are so happy today! What wonderful thing happened?'), or "Nummer 6 … Ich soll dir 100 DM leihen? Wozu denn?" ('Number 6 … I should lend you 100 German marks? What for?') (p. 45).

The picture stimulus of Part B allows the student to look at a picture sequence consisting of six black and white pictures for two minutes before the student gets another two minutes to "describe the scenes and relate the events in each of the six pictures" (p. 46).

Here, I will discuss only the rating guides for the speaking part, as none of the other parts of the ETS are of interest for this study. Appendix D, ETS Scoring Criteria[1], shows that the criteria for scoring directed responses is on a scale from 0 to 5:

- ranging from the lowest 0 - failure to understand the question; evading problem posed by giving an irrelevant answer; not attempted; total blank; English answer;
- to as high as 5 - excellent, demonstrates excellence; natural and meaningful responses; ease of expression; impressive range of vocabulary; no significant structural errors; no noticeable strain and/or stumbling; excellent pronunciation).

(see College Board, 1993, pp. 62-63; for further criteria see Appendix D)

---

[1] The reader can refer to the most current scoring criteria on the College Board/ETS website at: http://www.collegeboard.org/ap/german/frq99/index.html.

Additional guidelines are considered in cases of doubt. These include considerations of pronunciation to adjust the rating up or down, and sentence structures in which the use of basic or complex sentence structures can affect the rating either way (cf. p. 63). The complete list of criteria for picture sequences is found in Appendix D, but again, the rating scale ranges from the lowest possible:

- 0 - an irrelevant answer not addressing the picture sequence; answers in English; total blank;
- to as high as 5 - excellent – demonstrates excellence; strong control of and fluency in the language; scarcely any significant errors in structure; impressive range of vocabulary and idioms; excellent (near native) pronunciation.

  (see p. 64)

In addition, the student can gain an extra point for "exceptional fluency and/or a well-told story", but can also lose a point "for extremely poor pronunciation" (p. 64) if comprehension is hindered. Calculating each grade per student for the entire examination is described in detail in the 1992 examination booklet, although it is a rather lengthy procedure. The speaking part, in order to determine the reported grades, is weighted as approximately 25% of the total exam, even though the speaking part takes up only about twenty minutes out of the allotted 180 minutes in total. The remainder of the exam makes up 75% of the student's grade, with the listening comprehension part worth 25% and the rest (vocabulary, reading comprehension, grammar, paragraph completion, and composition) weighing 50%. It is important to note that the target group for this examination consists of students entering post-secondary education, who are required to perform in all four foreign language skills (reading, writing, listening, and speaking) in any further studies, with a strong emphasis on written work.

For the preliminary study that I conducted, the purpose of which was to gain insights into the students' German-speaking abilities by Grade Six (described in detail in the following chapter), I chose Part B of the ETS and its scoring criteria. The directed responses and picture stimuli criteria are easy to employ and student evaluation is quick. The criteria for scoring, however, proved to be inadequate for these students in Grade Six since the ETS rating scales do not supply adequate information about students' grammar and vocabulary. Nevertheless, the ETS guidelines provided a useful starting point.

3.3. TEST DEVELOPED FOR FRENCH TESTING IN CANADA

Dicks and Rehorick (1995) describe the Maritime Oral Communication Assessment Portfolio, MOCAP, which "is an evaluation package for French as a second language" (p. 135) for Grade Six and Nine students in a French core and immersion program. According to Dicks and Rehorick, the MOCAP involves an evaluation of students' oral abilities" focusing "upon the communication of messages, ideas, opinions and feelings," and furthermore, "the evaluation techniques can apply to any second/foreign language" (ibid.). However, the goal of this particular oral assessment tool is to evaluate students over a longer period of time, and not only to record the students' proficiency in the tested language at a specific moment in time.

The test tasks consist of a wide variety of separate tasks, which are made up of tasks involving "aural comprehension (listening) and oral production (speaking)" (p. 137) which are evaluated together, not separately. The MOCAP test tasks consist of seven parts:

1.) describing picture sequence (students talk about illustrations while looking at them);
2.) oral summary (giving an account in their own words after listening twice to a tape-recorded passage);
3.) forms (interview between two students where one student must fill out a form, e.g., an application for summer camp);
4.) questionnaire (involves two students and a form, similar to the above, but students have to add opinions and attitudes to factual information);
5.) information gap (students work together to obtain information; each student's information is incomplete without the other's);
6.) role plays (two students have two minutes to familiarize themselves with information and instructions); and
7.) discussion (four students at a round-table group discussion about a topic they familiarize themselves with within two minutes prior to the discussion).

The evaluation sheet used for each of these seven techniques can be found in Appendix E. In addition to the teacher's evaluation each student fills out a self-assessment questionnaire consisting of three parts. Part 1 asks the student about the actual language tasks and the student's ability to communicate messages and ideas. Part 2 requires the student to comment on specific strengths and areas that they need to work on, such as necessary vocabulary or expressions for a task, grammar, and correctness. Part 3 asks the student to globally evaluate his or her performance on an activity.

Dicks and Rehorick do not provide a time limit for the MOCAP tasks, but it is obvious from the seven parts described above that the length of the test encompasses several days, and in addition, is designed to evaluate students over a long period of time. As mentioned earlier, based on the time factor the MOCAP is unreasonable for the Grade Six teacher in my study, who requires an oral test for the Grade Six students that takes not much time to administer but offers a lot of information about the students' oral skills in German.

In this chapter, a variety of tests presently in use were examined and discussed, and some of the tests' advantages and disadvantages for the test reported in this thesis were illustrated. It was determined, that the ACTFL guidelines supply useful guidelines for the oral proficiency test introduced in this thesis. Whereas the OPI and the SOPI are unsuitable, the SOPA offers valuable evaluation criteria and the TOPT useful test material. The College Board and ETS criteria proved to be inadequate for the target group of the oral test in this thesis, and so is the MOCAP, the only Canadian test discussed in this chapter.

CHAPTER IV.

TEST DEVELOPMENT AND DESIGN

In this chapter I describe and discuss the preliminary study conducted with three Grade Six students who were orally tested in German, and the effects that the findings of this study have on the design of the first test version. Furthermore, the actual test is outlined and administered to three more students in a field testing situation. Based on the results of the field testing, all necessary modifications are discussed and applied to result in the polished and final version of the oral proficiency test, as well as its scoring and grading criteria.

4.0. PRELIMINARY STUDY

4.0.1. Methodology

In order to begin with the actual test design, I collected sample data from three Grade Six students in the German-English bilingual program approximately six months prior to developing the first test version. The goal of the preliminary study was to gather information about the oral language abilities of the Grade Six students in this particular elementary school in Alberta. The one male (M1) and two female students (F1, F2) and their parents consented to a short test which was conducted outside of the classroom in the homes of each participant, and the only persons present during the tape-recording were the participant and the researcher. Each test was conducted entirely in German, took approximately fifteen minutes, and contained four parts. The instructions and questions for the four parts are presented in Appendix F. Part I began with a short warm-up in German in which I asked the participants to describe briefly their family and their interests or hobbies. The warm-up serves as a connection, helping the student to switch from L1 to L2. Part II consisted of three questions relating to a variety of topics. These included questions as to what they would do if they won one million dollars, about holidays, and their best friend. In Part III, each student narrated a wordless picture book story, "Frog goes to dinner" (Mayer, 1977). The story consists of 22 black and white pictures showing the adventures of a frog in a restaurant. Mayer's wordless picture books have been used widely in monolingual and second language/bilingual research to investigate narrative strategies, which is why I used one of his books as part of my study. For Part IV, a short "cool-down" was administered, also in German, in which the students were asked what they thought about the questions (e.g., was it difficult to answer these questions in German?) and the picture book (e.g., did you like the story?). This final cool-down question serves as a closing, signalling the participants that the test is completed. All responses were rated according to the 1992 ETS Scoring Criteria (Appendix D), which were previously discussed in section 3.2. The tape recordings were analyzed to evaluate the different abilities, as well as possible similarities between the three students' oral abilities in these two very different tasks. Parts I and IV were considered in the evaluation as well, even though they are intentionally constructed to be metacognitive in nature. Most emphasis was allotted to Parts II and III.

Part II was designed to elicit free responses from the students, and the answers are also based on the student's comprehension of the question posed in German. The participants were able to reply freely to the topic presented in the foreign language. In this task, the student has certain flexibility in his or her response. Therefore, the student is able to decide on the direction the response takes, as well as the content that it entails. Part III, the picture book narration, differs greatly from Part II because the narrator is constrained by the objects and actions illustrated in the pictures. Moreover, the student's abilities to narrate a coherent and cohesive story (cf. Bamberg, 1997, Berman & Slobin, 1994) depend on his or her general storytelling skills and abilities, as well as the vocabulary and grammatical knowledge of the foreign language. The following section contains the responses of the three participants.

4.0.2. Results of the Preliminary Study

The production of the three participants were analyzed according to the following criteria: Fluency, pronunciation, sentence types, tenses, word order, self-repair, narrative style, language mixing (code-switching), and errors. These criteria emerged from initial examination of the data as categories: the researcher partially transcribed the utterances, and noted any additional characteristics, for example fluency and pronunciation, in form of comments. Furthermore, some differences and similarities between the three participants performing the tasks of answering questions and narrating a wordless picture book were noted. Table 1 shows that the three students did not perform with the same abilities when speaking German, even though they are in the same grade in this particular bilingual school. As planned, the preliminary study results illustrate what the students can do, even if certain criteria included in the list point out what the students can not do correctly in German at the time of testing.

Table 1: Analyses of language production

	M1	**F1**	**F2**
1. Fluency	+	+	+
2. Pronunciation	+	+	+
3. Sentence types (main & subordinate clauses)	+	+	+
4. Tenses (simple & compound)	+	+	+
5. Word order*	some English	some English	frequently English
6. Self-repair **	-	+	+
7. Narrative style ***	+	+	+
8. Language mixing (code-switching)	some (nouns)	frequently (nouns, verbs, adj., adverbs)	frequently (words & clauses)
9. Lexical retrieval difficulties	-	some	frequently
10. Errors:			
- case	few	many	many
- verb-particle, verb-prep.	few	many	many
- regularizing def. article	-	“die”	“den”
- congruence (subject–verb)	-	few	many
- ‘Germanized’ form	-	few	-
- G-E compounding	-	-	few
Evaluation	4-5	3-4	2-4

* M1 and F1 use mostly correct German word order.

** M1 does not need to correct his speech. Most of F2’s self-repairs are incorrect.

*** Individual differences: M1 offers interpretations, F1 mentions many details, F2 assigns a name.

The first criterion all three participants share when speaking German is the speech flow (the fluency of speech in German). All three participants are fluent according to Fillmore’s (1979) definition of fluency: the students are able to speak continuously, speak reasonably well about a range of topics, speak eloquently and thoughtfully, and use the language creatively (see section 2.4.). However, the students do not always speak perfectly fluently. Regardless of speaking continuously, several pauses were observed, but based on the students' language abilities, it is possible that these pauses are caused by language planning: planning *what* to say, and thinking less often about *how* to answer a question or describe a picture. In view of the fact that there is not enough evidence for language planning in the available data, it is not my intention to speculate further in this direction in the study reported here. The three participants do not differ in terms of the second criterion since they have native-like pronunciation when speaking German.

The next three criteria deal with grammatical structure. None of the participants had difficulties creating complex sentences that contain a main clause and a subordinate clause, even if the clauses contain grammatical, structural (syntactic), and/or morphological errors. In addition, all the students use simple (e.g., present tense, simple past) as well as compound tenses (e.g., present perfect) during the question period and the narrative. While all three participants apply English word order in some cases, F2 produces more sentences that are influenced by English syntactic rules than M1 and F1. Most of these errors are found in the incorrect position of the verb, as for instance in "... *dann ich *würde ...*" ('then I would') instead of "… dann würde ich …".

Both F1 and F2 self-repair single words as well as complete clauses, and in addition to that, F1 frequently asks questions in order to confirm that she understood the questions correctly. Many of F1's self-repairs are correct. F2, on the other, attempts to self-repair many utterances as well, yet most of her corrections are incorrect. The participants possess linguistic awareness and this indicates that even though F2 is able to detect certain errors, she is not able to repair her errors satisfactorily. Partly due to this observation, F2 appears to be less proficient in German-speaking than M1 and F2.

There are differences in the students' story-telling skills, abilities, and preferences. Part III (see Appendix F) of the preliminary study consists of narrating the wordless picture book. Each participant's story is characterized by different phenomena: M1 offers interpretations to the pictures he sees and the faces he observes, e.g., "*der ist sauer*" ('he is upset'), "... *lacht und denkt das ist lustig*" ('... laughs and thinks that is funny'), and "*er ist froh, dass er sein' Frosch zurück hat*" ('he is glad that he got his frog back'). F1, on the other hand, mentions many details she observes in the pictures that are not relevant to the main actions portrayed, such as details seen in the background of the pictures. And finally, F2 is the only student who assigns a name to the boy in the picture book. Clearly, these individual differences in telling a story support the general phenomena present in narration (see Berman & Slobin, 1994). Individual differences are not specific to one language and the participants show this phenomenon when speaking a foreign language, German in this case.

One observation that is not included in the table above and that three participants share, is their difficulty with the task of telling the picture story. Their speech flow is interrupted as they tell the story, they speak much more slowly than in Parts I and II (answering questions), and create many pauses. The possible causes for this phenomenon are discussed in section 4.0.3.

Code-mixing with English is very common among the three participants, even if the types and tokens of switching differ. M1 uses relatively little English vocabulary, and all of the English terms are nouns, such as *motorboat, bodyguard, big-screen TV, jet, Universal Studios, cruise ship, limousine,* and *first class hotel*, and he pronounces these words according to English phonological rules. Most of these words would, however, be used in German the same way, such as *Jet* and *Universal Studio*, although M1 may not be aware of this. I believe there are at least three possible explanations for the use of these words in English: first of all, the students in this school are accustomed to mixing English words into their German utterances to a varying degree. The reason for mixing lies in the inevitable fact that these children are learning a minority language in an English-speaking culture. In addition, every teacher in the school mixes English into her German, some more, some less. All of the teachers in this bilingual program were either already born in Canada, or have lived the majority of their life here. Consequently, their German is strongly influenced by English, and even though they try to avoid English during certain times in class, it is unavoidable for the dominant English not to be used. A second reason is, that code-switching is a language strategy bilingual speakers employ, a phenomenon discussed in a variety of linguistic literature (i.e., Baetens Beardsmore, 1986). Finally, many of the words M1 uses are often used by native German speakers, as for example, *Jet*, *Universal Studio*, *First Class Hotel* or even *Limousine*. Whereas the first three words are of English origin, the latter one is originally a French word and native German-speakers pronounce it [li-mu-'si:ne].

F1 uses a variety of English words as well, as for example *"ich könnte nicht eine *decision machen"* ('I couldn't make a decision'), but F1 mixes also some verbs, adjectives, and adverbs, however, not as frequently as participant F2. F2 mixes not only single words, but also complete clauses into her German speaking, such as *"we never fight"*.

The next point in Table 1, lexical retrieval difficulties (or automaticity), is yet another phenomenon of linguistic awareness. Automaticity refers to the language learner's ability to retrieve lexical items. The less advanced a learner, the more time does it take to retrieve lexical items. A more advanced learner takes less time to retrieve lexical items. M1 shows no retrieval problems at any time of the recording session, but F1 and F2 have difficulties retrieving vocabulary. F1 had difficulties remembering the German word *"traurig"* ('sad') at a point during her narrative and uses the English equivalent instead. When she remembers the German term later on, she employs it correctly. F2, on the other hand, had difficulties remembering many words (nouns, verbs, adjectives, and adverbs), even words as familiar to her as *"Familie"* ('family'). This observation indicates that M1 is orally more proficient than F1 and F2, but F1 is more proficient than F2.

The last criterion listed in Table 1 is based on a partial error analysis. The types of errors in Table 1 were categorized according to the detected errors that were transcribed from the recordings. The following error categories are distinguished: case, verb-particle and/or verb-preposition combinations, regularizing definite articles, congruence (subject-verb agreement), 'Germanized' forms (Clyne, 1985, refers to these forms as grammatically or phonologically integrated lexical transfers), and German-English compounding. The three participants perform differently in most of these categories, however, there are strong similarities in creating case errors and verb-particle and verb-preposition errors. M1, for example, uses the nominative case instead of the accusative in a few instances. Both F1 and F2 use all four cases (nominative, accusative, dative, and genitive) incorrectly more frequently, e.g., the incorrect accusative pronoun after the preposition *"mit *mich"* ('with me') instead of the correct dative pronoun 'mit mir'. All three students produce errors in the misplacement and production of verb-particle and verb-preposition constructions, such as *"hat gedacht *über"* (present perfect; 'has thought about') instead of 'gedacht an'. Many of the incorrectly produced prepositions or particles would be correct in the English equivalent.

Generally, M1 does not have problems with as many grammar categories as F1 and F2 have. F1 regularizes the definite article *'die'* ('the'; nominative and accusative form of first person singular feminine and plural) and uses it in most instances when an article is called for, however, many of its applications are incorrect. F2 also regularizes a definite article, but it is *'den'* ('the'; accusative form of first person singular masculine and dative plural) e.g., *"der Mann und den Frau"* ('the man and the woman') instead of 'der Mann und die Frau', or *"*den Frosch *looked confused"* ('the frog looks confused') instead of 'der Frosch …'.

Congruence errors occur a few times with F1, e.g., *"der Junge *hast gesehen"* ('the boy *have seen'; '*hast*' is second person singular) instead of 'hat' (third person singular), and also many times with F2, such as in *"ich *hat zwei"* ('I *has two') instead of 'habe' (first person singular). Only F1 uses a 'Germanized' verb form, consisting of the English stem and the German third person singular –t ending in *"*wave-*t"*, instead of 'winkt' ('waves'). F2, on the other hand, is the only student to combine a compound in English and German, as in *"Eier-*hunt"* ('egg hunt') instead of *Eiersuche*, however, this happens only in one instance.

I evaluated the three participants' test performances according to the ETS Scoring Criteria (see Appendix D for complete criteria). It was impossible to score just one criterion definition for each participant, because F2, for instance, has good pronunciation (score 4) in both tasks, but illustrates 'some major errors in structure' (score 3), as well as 'use of some anglicism [sic.]' (score 2). The same combination problem is found with M1 and F2 and resulted in a combination score for all three participants. M1 scores between a 4 (good) and a 5 (excellent) in both tasks, namely Part A (Scoring directed responses) and Part B (Scoring picture sequence). F1 was also evaluated for both tasks between a 3 (acceptable) and a 4 (good), and F2's evaluation ranged for both tasks from a 2 (weak) to a 4 (good). The problem with the evaluations is found in the definitions of the ETS criteria. These results illustrate the difficulty of evaluating the participants accurately. Moreover, the range of scores demonstrates the weakness of the ETS Scoring Criteria. The evaluation range for each participant supports the inadequacy of this test for the rating scale of the oral proficiency test that is required for the Grade Six students in this bilingual elementary school. Nevertheless, the preliminary study offers helpful information about the students and their abilities in speaking German.

4.0.3. Summary, Discussion, and Implication for the Field Testing

There was a noticeable difference between the students' free responses versus the picture book narration with respect to the speakers' fluency. The free responses appear to be an easier task for the participants, and the answers are delivered with fluency and ease. The narration of the wordless picture story on the other hand, slows down the responses of the speakers, and the pauses that are created appear to be a sign of dysfluency, maybe even of lower proficiency. The fact is, the task of telling a wordless picture story involves a variety of skills and abilities, namely looking at the picture closely to combine the visual input with the participant's knowledge of the world. The speaker must connect persons, objects, and actions with one another, as well as know the vocabulary and grammar of the non-native language to describe what is seen in the picture. Moreover, the narration of a story entails discourse strategies such as coherence and cohesion devices (see Bamberg, 1997, Berman & Slobin, 1994), which are necessary to narrate a story.

In sum, even though the students in Grade Six produce a variety of errors in several grammatical categories, all three students show that they are able to perform many different tasks in German. They are able to speak fluently, with native-like pronunciation, and they can produce a great variety of sentence types, such as main clauses and subordinate clauses, and relative clauses. They can use a variety of tenses, namely simple tenses (e.g., present tense, simple past) and compound tenses (e.g., present prefect), which allow them to speak about events in the past, present, and future time. Two out of the three participants are able to use German word order in most cases, but one participant, F2, uses English word order frequently.

The implications of the preliminary study results suggest that the children's grammatical errors and the quantity of mixing English into German utterances must be treated differently than in existing oral tests. None of the existing tests discussed in Chapter 3 take the type and amount of code-mixing into account. In order to avoid as much mixing as possible, the students should be instructed to attempt to speak only German, something that was not stressed in the instructions during the preliminary study. Even though the students in the preliminary study were instructed to speak German, they were not explicitly instructed to avoid English, if possible. If these instructions are given to a student before the test and the student mixes German with English, the amount of mixing could be a reflection of the student's oral proficiency in German. Furthermore, the children's pronunciation should only be considered in the evaluation if it hinders the listener's comprehension, since all three children showed native-like pronunciation.

Based on the criteria list presented in Table 1, it appears that the evaluation criteria for the oral proficiency test proposed in this thesis should take into consideration the following grammatical areas:

- word order;
- mixing (quantity and quality);
- case;
- verb-particles, verb-prepositions;
- possible regularizations;
- congruence (subject-verb agreement);
- possible 'Germanization' and German-English noun compounding.

Word order, case, verb-particles and verb-prepositions, possible regularizations, and congruence are aspects of accuracy, and are considered to be aspects of grammar and vocabulary. Code-mixing, possible 'Germanization', and German-English noun compounding are all features of code-switching and transfer, and even if code-switching is discussed in linguistic literature extensively (e.g., Baetens Beardsmore, 1986), it is not accounted for in any of the previously discussed oral tests.

Additional criteria to consider in the design of an oral proficiency test could include speaker fluency, self-repair, and retrieval (automaticity) difficulties. Speaker fluency is defined and discussed in Chapter 2, and L2 self-repair "studies are very few in number" (Van Hest, Poulisse & Bongaerts, 1997, p. 86). Self-repair and automaticity will not be discussed further in the study as they are not the focus of the oral proficiency test in this thesis.

The majority of the criteria listed above include aspects of fluency and accuracy, but not aspects such as content (topic), context (situation), discourse, and narrative abilities. In order to evaluate the students' oral proficiency successfully, these aspects should be included in the evaluation and rating scale. The nine-level SOPA rating scale which includes fluency and accuracy (vocabulary and grammar), is certainly of interest if I compare the SOPA criteria with the areas identified in the preliminary study.

4.1. FIELD TESTING PREPARATION AND FIRST TEST DESIGN

4.1.1. Test Preparation

The Grade Six teacher of the German-English bilingual elementary school and I, discussed test design, question types, question content, evaluation procedures, grading criteria, and grading scale. Edmonton Public Schools (1991) define the key learning outcomes (goals) for 9- and 12-year-old bilinguals:

> 1. Students have the ability and desire to read, listen, and view for a variety of purposes.
> 2. Students express thoughts and feelings for a variety of purposes in a variety of situations.
> 3. Students responds to literature and other art forms.
> 4. Students appreciate and participate in the culture.
> 5. Students use language as a medium for cognitive, affective, and social learning.
> 6. Students have the knowledge of language and the skills necessary to interpret and express ideas. (pp. B1-B8)

Based on these expected outcomes, the oral proficiency test for the Grade Six students proposed in this study should include a variety of situations and topics, including literature and art forms, as well as German culture. Furthermore, knowledge of the language must be tested, as well as the students' skills and abilities to speak German. An important issue discussed and decided on was the grading system, and how to match the teacher's expectations of the Grade Six students in this bilingual program with the SOPA criteria and the school district's grading system. The school district's grading system, which describes the student's performance, is divided into the following grades and descriptors:

Letter Grade	Percentage Grades	Descriptors
A	80-100	Work meets standard of excellence
B	65-79	Work exceeds acceptable standard
C	50-64	Work meets acceptable standard
D	0-49	Work does not meet acceptable standard

(Edmonton Public School Board, 1995)

Even though a percentage grade is supplied to correspond to the letter grades, public elementary schools in Edmonton do not record percentages on the progress report card. Progress report cards vary from school to school, but students receive letter grades, and in some cases descriptors, as well as a detailed progress description for every subject matter. Based on the school system's grading, the evaluation criteria for the different SOPA categories and levels, the students will also receive letter grades ranging from A to D in the oral proficiency test.

After reviewing all of the existing oral tests discussed in Chapter 3, as well as the preliminary study results and observations, the Grade Six teacher and I agreed that a comprehensive oral test, such as a combination of the SOPA and the TOPT, fulfills almost all of these expectations. Consequently, we decided that the oral proficiency test appropriate for the students in this bilingual program will:

- consist of four parts: a brief warm-up, a short-answer section, a picture book narration, and a brief cool-down (entirely in German);
- be tape-recorded during the trial run, as well as the actual test;
- follow the existing SOPA criteria (with some additions) using the student performance criteria used in Edmonton's public elementary schools;
- use the TOPT speaking tasks, but reword all of the possible questions to be appropriate for Grade Six students, as well as change and add questions according to topics, cultural content, and language knowledge.

We then decided that the Grade Six teacher would use the test with three students in order to verify the time factor (how long it takes to record each participant), the SOPA criteria, and the reworded TOPT speaking tasks. The teacher tape-recorded, evaluated, and graded each student according to the guidelines and criteria decided on. In addition, a second Grade Six teacher from the same school was enlisted to these three tape-recordings and evaluate and grade the students as well. Finally, I evaluated and graded the students and all three evaluations and grading results were compared. The goal was for all three evaluators to reach the same evaluation and grading results in order to support that the test is considered appropriate and reliable (see discussion in section 2.2., Finocchiaro & Sako, 1983).

4.1.2. Test Design, Description, and Procedure

Following all of the considerations discussed in the section above, as well as the previously discussed measures of which criteria should be included in the test and its evaluation, the actual test was designed. The Grade Six teacher matched the SOPA criteria with expectations for the students in Grade Six in German and as a result, Table 2 illustrates a condensed version of the adapted SOPA categories and letter grades as decided on by the teacher in compliance with the key learning goals (see section 4.1.1.) of Edmonton Public Schools (1991).

The entire version of the Rating and Letter Grading Scale for Bilingual Grade Six, Appendix G, clarifies what each category and level entails. As illustrated in Table 2, any performance at the Novice level is considered a failing Grade D (work does not meet acceptable standard). A student who scores at the Jr. Intermediate Low level in fluency or vocabulary also receives the failing Grade D. All students are expected to at least meet the acceptable standard, a Grade C. Any performance in speaking German beyond the Intermediate level exceeds acceptable standards, and the student will either receive a Grade B or A respectively. In order to receive a Grade A, the Grade Six students must perform at the Jr. Advanced level, which is the level considered to meet the standard of excellence.

Table 2: Letter grading scale adapted to the SOPA

	Novice (N)	**Jr. Intermed. Low (IL)**	**Jr. Intermed. Mid (IM)**	**Jr. Intermed. High (IH)**	**Jr. Advanced (Ad)**	**Jr. Advanced High (AH)**	**Superior (S)**
Listening Comprehension	D	***C***	***B-***	***B+***	***A***	A+	A+
Fluency	D	D	***C***	***B***	***A***	A+	A+
Vocabulary	D	D	***C***	***B***	***A-***	***A+***	A+
Grammar	D	***C-***	***C+***	***B***	***A-***	***A+***	A+

Note: highlighted letter grades refer to what is expected of the students in Grade Six in order to pass. Performances at the Novice level are considered a fail, performances at the Jr. Advanced High and Superior level surpass expectations.

4.1.3. First Version of the Test

After deciding on using the SOPA criteria, the TOPT speaking tasks were reworded to suit the target group of this oral proficiency test, the Grade Six children. Appendix H shows the Preliminary Test design which includes all four sections of the oral proficiency test (English translations are provided for the German questions and instructions), and is followed by a Grammar Checklist (Appendix I) and Additional Evaluation Criteria (Appendix J). The latter two serve the teacher as additional tools for the test evaluation and grading procedure. The four sections of the oral proficiency test consist of:

- Section I: Warm-up
- Section II: Short answers

    A. Present tense: repetitive; descriptive (easy; more difficult; complex situations, concrete)

B. Past tense (simple past; present perfect): descriptive (easy; more difficult)

C. Future tense: descriptive (easy; more difficult)

D. all tenses possible: comparative (easy; more difficult)

E. Hypothetical, "what if" situations

- Section III: Picture book narration
- Section IV: Cool-down

Section I consists of a short warm-up question which allows the participants to start speaking and thinking in German. The answer requires the participant to state his or her name and age, and a short statement of the number of brothers and sisters the participant has, where the participant lives, and what his or her favourite food is. Section I is followed by Section II, one of the two principal parts of the test. Section II is the short answer speaking task which includes many reworded questions of the TOPT, as well as additional questions and situations. The section is divided and categorized according to the type of answers required from the participants. The type of answers required are divided into five categories, from A to E. The questions in parts A, B, C, and D prompt the participants to use a particular grammatical tense in his or her answer: A = present tense, B = past tense (simple past or present perfect), C = future (expressed usually in the form of present tense), or D = a possible choice or combination of tenses. The last major task, E, consists of a question about a hypothetical, "what if" situation, in which the participant is required to play a role in order to answer the question.

Most of the questions in parts A to D are divided into task types (repetitive, descriptive, or comparative), and into level of difficulty (easy or more difficult). The easier questions and answers will be used at the beginning or in the middle of the Grade Six school year. The more difficult questions can be used at the end of the school year.

The repetitive tasks in part A (present tense) consist of questions that depict a repetitive action, for example, how to loan a book from the school library or to talk about a typical school day. The descriptive tasks in part A are divided into easy and more difficult questions. The easier questions refer to describing oneself, likes and dislikes, one person or more, a place, an object, or an action. The more difficult questions consist of the description of a bicycle or car, the participant's house, apartment, or room, a particular piece of art work, the city of Edmonton, holidays such as Christmas, the country of China, or Alberta's natural resources. An additional complex task consists of a concrete situation in which the participants have to imagine themselves, and to describe how they will react in this situation. In one example, the participant is put into the situation of being approached by a friend or fellow student who wants to sell drugs to the participant. The participant is asked what he or she is going to say and do.

The questions of part B prompt the participant to use the past tense in his or her answer, either the simple past or the present perfect. The answers are of descriptive nature and the participants are asked to describe an event or activity that took place in the past. The questions are relatively easy, but some are more difficult. In one of the more difficult question, the participant must describe what he or she knows about how the ancient Greeks used to live.

The questions in part C prompt the participants to use the future tense, which in German is commonly expressed in the form of the present tense form. However, a conjugated form of '*werden*' ('will') + infinitive can also be used to express future events or actions. In the easier questions, the participants are asked about plans for the coming weekend or holidays, and the more difficult questions ask the participants about plans for the future as an adult, a future holiday destination, or how to give somebody detailed directions.

The questions in part D allow the participants to choose between different tenses in order to compare a variety of things. In the easier questions, the participants compare cultural differences between Canada and Germany, for example Christmas traditions, or Halloween and Germany's Carnival ('*Fasching*'). The more difficult questions demand the participants' ability to compare the lives of the ancient Greeks or the Chinese with life in Canada.

In addition to using particular tenses for a specific task, hypothetical, "what if" situations are given in part E in which the participants should use conditional forms in their answers. The questions include hypothetical situations, such as what to do after winning one million dollars, having to move to China or Germany and what to do there, trying to get out of a tricky situation, meeting ones hero, or living in the past with the ancient Greeks. During the test, the teacher should, if possible, ask one question from each subsection, A, B, C, D, and E. This enables the teacher to test all necessary tenses, as well as specific topics and vocabulary, specific grammatical areas, and cultural knowledge.

After Section II, the participants look at the wordless picture book "Frog goes to dinner," (Mayer, 1977) and narrate the story. The picture book narration is followed by a short cool-down question, Section IV, in which the participants either comment on the story and what they thought of it, how they liked it, or how they feel after the test.

The test proceeds as follows: before the test, the Grade Six teacher chooses five questions from Section II. The teacher will be meeting with only one participant at a time for the test. The test session is tape-recorded and the teacher listens to the tape at a later point to evaluate each participant accurately. It is preferred that the teacher does not make any notes during the recording but immediately after the test, if necessary. This allows the teacher to concentrate on the questions and answers, get involved with the participants, and most important, the teacher will not make the participants more nervous than they most likely already are in a test situation in which they are tape-recorded. That way, the participants will not worry about what the teacher may be writing about their German-speaking abilities. In addition to the actual test sections, the teacher will be using two additional 'tools' to aid the evaluation procedure. The first tool is a Grammar Checklist (see Appendix I) and the second one consists of two Additional Evaluation Criteria (see Appendix J) necessary to evaluate the students in this bilingual Grade Six appropriately.

The Grammar Checklist includes structures the participants are expected or not expected to know or use. This list allows the teacher to concentrate on some grammatical features but ignore others respectively during the evaluation. On the one hand, the participants are expected to know and use the following grammatical features correctly:

Tenses

- Present Tense: most verbs, modals, '*sein*' (to be) and '*haben*' (to have);
- Simple Past: most verbs, modals, '*sein*' and '*haben*';
- Present Perfect: most verbs, '*sein*' and '*haben*';
- Future: present tense; '*werden*' (will) + infinitive.

Word order

- verb position in main clauses (time-manner-place);
- verb position in subordinate clauses;
- subordinate clauses with the conjunctions '*dass*' (that) and '*weil*' (because);
- imperatives (verb in initial position, singular and plural forms);
- subject, object, indirect object (by the end of the school year);
- relative clauses in the nominative case (by the end of the school year).

Other grammatical areas

- nouns (singular and plural formations);
- formal address with '*Sie*' (formal 'you');
- imperative forms;
- prepositions (accusative and dative);
- adjectives (comparatives, superlatives);
- possessive pronouns (nominative);
- adjective endings.

On the other hand, the participants are not expected to know and use certain grammatical forms correctly. Students do not receive credit for forms that are not expected to be known and used because in most cases, the students producing these forms correctly are native German-speakers. If the participants produce grammatical errors within any of these areas, the teacher will ignore them, and the participants' oral proficiency evaluation is not influenced by these errors:

Not expected

- modals in present perfect;
- past perfect;
- future perfect;
- subjunctive mood;
- genitive;
- possessive pronouns (accusative and dative).

This grammar checklist is not the only tool available to the teacher; it simply serves as a reminder during the evaluation. Clearly, grammar is not the only criterion to consider during the test, since grammatical accuracy alone does not reflect oral language proficiency. Nevertheless, grammar is one criterion in the SOPA and will be evaluated and graded. In addition to the checklist, the teacher will concentrate on many other aspects in the participant's German speaking, some of which are categories included in the SOPA:

- listening comprehension;

- fluency;
- vocabulary (often topic specific, as necessary for talking about the ancient Greeks or natural resources).

The second tool the teacher will use includes two criteria that the SOPA does not list as individual categories, and are not included in the SOPA evaluation:

- cultural knowledge about Germany (e.g., when comparing German and Canadian traditions); and
- mixing English into German (code-switching).

(see Appendix J for detailed description)

Whereas listening comprehension, fluency, vocabulary, and grammar are included in the SOPA criteria, the latter two, cultural knowledge, and mixing with English, are not considered. After the preliminary study, the researcher and the Grade Six teacher decided, that these latter two criteria must be included in the oral proficiency test for the students of this bilingual school. Cultural knowledge is an important feature of the German-English bilingual program in which the students learn about German traditions and how life in Germany compares to life in Canada. Cultural knowledge is not discussed in any of the reviewed oral tests, and even though the term 'culture' is referred to in the ACTFL guidelines' (1999) introduction, ACTFL does not specify what culture means and how it becomes a part of the language learner's knowledge. It is also not clear why the SOPA, and other oral proficiency tests, do not incorporate reversion to L1 (code-mixing or code-switching) in their evaluation criteria. The ACTFL Proficiency

Guidelines – Speaking (Appendix A) refer to the dominant language, and its effect on the second/foreign language, however, ACTFL's brief mention is not sufficient for the target group discussed in this study. There is no mention of code-mixing at all in the following levels: Superior, Advanced High, Intermediate Mid, Novice Mid, and Novice Low. The only mention of L1 are in the following levels:

- Advanced Mid: Dominant language discourse structures tend to recede, although discourse may still reflect the oral paragraph structure of their own language rather than that of the target language.
- Advanced Low: Structure of the dominant language is still evident in the use of false cognates, literal translations, or oral paragraph structure of the speaker's own language rather than that of the target language;
- Intermediate High: the dominant language is still evident (e.g., use of code-switching, false cognates, literal translations, etc.);
- Intermediate Low: Their pronunciation, vocabulary, and syntax are strongly influenced by their first language;
- Novice High: These speakers' first language may strongly influence their pronunciation, as well as their vocabulary and syntax when they attempt to personalize their utterances.

(see Appendix A)

The results from the preliminary study showed that the ACTFL mention of code-mixing do not include what was observed during the preliminary study, namely, mixing at possibly every level. Consequently, in addition to the SOPA criteria and the grammar checklist, an additional list was composed that considers these two areas: cultural knowledge and code-mixing. Cultural knowledge about Germany is an area the students are taught specifically in class, and with it they learn domain-specific vocabulary that they are expected to know and use. Therefore, this particular criterion is included in the students' evaluation reflecting a set of acquired vocabulary, vocabulary growth, and the knowledge of more complex topics of German cultural content.

Clyne's (1985) study with children in two very different bilingual language programs in Australia pointed to some interesting code-switching situations in the two schools. In the bilingual program of school 1, only German is spoken inside and outside of the classroom, and formal grammar is not taught. In the second language program of school 2, German is used two half hours per week, and emphasis is on grammar. Clyne distinguishes between six types of code-switching and transfer phenomena:

- Grammatically integrated lexical transfer (e.g., ich runne, gewalkt);
- Phonologically integrated lexical transfer (e.g., Kupf, "cup");
- Semantic transfers (e.g., Mädchenfreund, "girl friend");
- Mixed idioms (e.g., Ist das *right*? Was sprichst du about?);
- Neologismus (e.g., Händewächer, face-washer, wash-cloth);
- Code-switching anticipating a trigger word (e.g., Vater steht bei dem Fleisch und *drinkt beer* (Bier)).

(pp. 204-205)

The children in the bilingual class (school 1) used all of these code-switching techniques, whereas the children from the structured second language class (school 2) were unable to incorporate English materials and usually fell silent instead. Clyne explains that the students in the bilingual program are function-oriented, with the purpose to communicate. The children in the second language program are structure-oriented with the purpose to keep the two languages apart (see pp. 207-208).

Code-mixing and transfer are common phenomena among the bilingual students in this study as well. These Grade Six students possess the same metalinguistic awareness as the students in the bilingual school in Clyne's study and they employ the same strategies of switching and transferring. Following Clyne's distinction, the children in the elementary school in Alberta are function-oriented. As observed during the preliminary study and confirmed by the teacher, all the students mix English into their German, however, the type and amount of mixing varies greatly from student to student within one Grade. I hypothesize that there may be a correlation between type and amount of mixing and the student's overall proficiency level, a hypothesis which cannot be answered here since the sample is too small. The preliminary study showed that the speaker who was considered the most proficient one, M1, mixed a small number of English words (nouns only) into his German, and many of these words are familiar to native German speakers too. M1 had no comprehension problems, was a fluent speaker of German, used a variety of vocabulary (some topic specific), and produced the least grammatical errors. It seems that F1 and F2, who were considered to be less proficient German speakers than M1, also mixed more English into their German. F1 was judged a more proficient German speaker than F2, and mixed less with English (nouns, verbs, adjectives, and adverbs) than F2. F2 was evaluated the least proficient German speaker among the three participants, and mixed single English words, as well as clauses, and even complete sentences into her

German. Further investigation into the possible correlation between mixing L1 into L2 goes beyond the scope of this paper.

4.2. FIELD TESTING

4.2.1. Methodology

The Grade Six teacher chose three students as students (S1, S2, and S3) to try the first version of the test. The parents of each student consented to the test participation of their child. The teacher tape-recorded each student separately, and followed the test procedure as intended: one brief warm-up question, five questions from the speaking task, the picture book narration, and one brief cool-down question. The teacher did not ask any questions concerning cultural knowledge about Germany in this field testing, however, she asked easier questions, as well as some more difficult questions (e.g., comparing China to Canada). The teacher did not ask all three students the same five questions for two reasons. First, the students could talk to one another about the test and prepare the answers, and secondly, the field testing of the first test version is intended to try five different questions with each student to test the questions' applicability and appropriateness. Therefore, it is considered necessary to ask many different questions. After the test, the teacher listened to the tape at a time convenient for her schedule, and evaluated the students according to the SOPA guidelines and the Additional Evaluation Criteria. Each student was assigned a letter grade for listening comprehension, fluency, vocabulary, grammar, and mixing with German.

After the completion of the evaluation, the teacher passed the tape and the evaluation criteria lists to a colleague who also teaches the bilingual Grade Six class this year, who listened to the tape and evaluated each of the three students accordingly. This teacher was not aware of the grades the first teacher assigned. Finally, I received the tape and evaluated the three students. I did not know the grades the two teachers had assigned. After my evaluation, I compared the three evaluations and grades we had assigned to each student.

4.2.2. Evaluation Results and Observations

Table 3 presents the inter-rater consistency and illustrates how the three evaluators E1, E2, and E3 (two teachers and I), evaluated and graded the three students speaking German. Even though the three evaluators do not completely agree in each category, the letter grade distribution for each individual criterion, as well as the overall final mark each student received, illustrate that the three evaluations are very similar. It must be noted that each evaluator listened to the tape once. The evaluators disagree only in minor level distinction which is to be expected from any evaluation. However, in future tests and evaluations, the evaluators will listen also only once to the tape, and therefore, the observed results are satisfying and serve their purpose: an overall test result for each student, based on evaluator agreement. The crucial result is, that the three evaluators scored the students almost identically, and mostly within one ACTFL level (e.g., Jr. Intermediate).

To begin with, we will look at the results each student obtained, and how the three evaluators assessed and graded the oral productions. The four categories, listening comprehension, fluency, vocabulary, and grammar, confirm a rather unison evaluation of the three students. I will discuss the fifth category, code-mixing, later in this section.

Table 3: Evaluation results and grades

	S1			**S2**			**S3**		
	E1	**E2**	**E3**	**E1**	**E2**	**E3**	**E1**	**E2**	**E3**
Listening Comp.	B+ *IH*	B- *IM*	B+ *IH*	A *Ad*	A *Ad*	A *Ad*	A *Ad*	A *Ad*	B+ *IH*
Fluency	C *IM*	C *IM*	C *IM*	A *Ad*	B *IH*	A *Ad*	B *IH*	B *IH*	C *IM*
Vocabulary	C *IM*	C *IM*	C *IM*	A+ *AH*	A- *Ad*	A+ *AH*	B *IH*	B *IH*	B *IH*
Grammar	C *IM*	C *IM*	C *IM*	A- *Ad*	A+ *AH*	B *IH*	B *IH*	B *IH*	C+ *IM*
Mixing English	D	*	D	A	*	A	B	*	A
Final mark	C	C	C	A	A	A	B	B	B

*** Due to researcher's error, E2 did not receive the additional evaluation criteria list and the category stays blank.**

S1 was evaluated to be orally proficient at the Jr. Intermediate Mid level ten times, and at a Jr. Intermediate High level twice. These latter two ratings at a higher level are in the listening comprehension category, and the actual letter grade difference is minimal: a B+ instead of a B-.

S2 was evaluated two times at the Jr. Intermediate High level (E2 for fluency, E3 for grammar) and three times at the Jr. Advanced High level (E1 and E3 for vocabulary, E2 for grammar). All other evaluations assessed the student at the Jr. Advanced level.

S3 was evaluated two times at the Jr. Intermediate Mid level (E3 for fluency and grammar), twice at the Jr. Advanced level (E1 and E2 for listening comprehension), and eight times at the Jr. Intermediate High level.

These results support the goal of field testing the first test version that the SOPA evaluation criteria prove effective as a tool for test procedure, evaluations, and results for the Grade Six students of this bilingual elementary school.

It is necessary to look at the evaluation disagreements closely in order to decide, if the different evaluation results stem from the test design or the evaluation criteria, or are caused by another source. The first instance of disagreement is observed with S2 and the evaluation of fluency. E2 evaluated S2 at the Jr. Intermediate High level which translates into a letter grade B, yet, E1 and E3 evaluated S2 at the Advanced level, resulting in a letter grade A. The second instance of disagreement occurs with the same student, S2, but shows disagreement in the evaluation of the student's grammar. E3 evaluated S2 at a lower level, Jr. Intermediate High, whereas E1 and E2 evaluated the student at a Jr. Advanced, and Jr. Advanced High level respectively, the overall final mark for student S2 is not affected. Even if only two of the three evaluators had evaluated and graded the student, the final mark still would have resulted in a letter grade A. The third instance of disagreement is found in the evaluation of the listening comprehension of S3, in which case E3 evaluated S3 at a lower Jr. Intermediate High level, instead of the higher level E1 and E2 had chosen, the Jr. Advanced level. Again, the overall final mark still results in a letter grade B for all three evaluators.

To summarize, the three evaluators judged S1 to be orally proficient at the Jr. Intermediate Mid level, resulting in a letter grade C, S2 as orally proficient at the Jr. Advanced level, resulting in a letter grade A, and S3 at a Jr. Intermediate High level, resulting in a letter grade B.

The additional criterion, code-mixing with English, does not influence the overall result: the majority of evaluations are in agreement, and the students' final marks were not situated between two letter grades, in which case the code-mixing mark can decide the final mark. Due to researcher error, E2 did not receive the additional evaluation criteria list, and consequently, could not evaluate this portion of the test. S1 received a D from E1 and E3, because the student's code-mixing included clauses and complete sentences. The student did not appear to attempt to speak only German. S2, on the other hand, received an A from both evaluators. Even though S2 used some English nouns in her German utterances, the amount of mixing was minimal, and the student made a strong attempt to speak only German. S3 received a B from E1 and an A from E3. Whereas E1 believed to have heard S3 use English verbs, adjectives, and adverbs in addition to some nouns, E3 did not hear the use of other English words than nouns, and considers the student to illustrate a strong attempt to speak only German. The different evaluation results do not influence the final mark of the student, a B.

One particular discussion point of the evaluations and final marks that Table 3 does not make explicit refers to the observations of amount of code-mixing of S1. The evaluators E1 and E3 commented on the difficulties of evaluating this student appropriately. Even though the student's overall evaluation was considered at a Jr. Intermediate High level, the type and amount of code-mixing and was so extensive, that it influenced the evaluation significantly. S1 appeared to have little or no problems comprehending the teacher's questions, however, the answers S1 produced contained more English than German, and due to this lack of German data, it was difficult to evaluate vocabulary and grammar. However, the apparent ease of mixing the two languages, as well as the native-like pronunciation of the German words and clauses gave the impression that S1 fits into the category of a Jr. Intermediate level speaker. Furthermore, E1 and E2, who teach S1 in class, know of the student's actual language abilities, even if S1 did not demonstrate these abilities in the oral test. Both E1 and E3 commented on this observation. E3 is also aware of these abilities, and it seems the evaluators were influenced by their knowledge about these abilities, which resulted in a letter grade C, a passing grade. After discussing this case and the amount of code-mixing observed here, the Grade Six teacher declared, that in her opinion, any student who does not attempt to speak only German will receive an overall failing letter grade D. Therefore, the additional evaluation criteria listed in Appendix J will not only serve to decide on in-between grades, but can lead to an overall failing grade, even if a student appears to perform at an overall passing level. In sum, the evaluation results and grades tabulated in Table 3 establish strong support for the test's practicality and scoring reliability, and after some changes and fine-tuning, the final test version is established.

4.3. MODIFICATIONS TO THE FIRST TEST VERSION

The Grade Six teacher and I discussed the field testing of the first version and its results in detail and based on the observations and discussion, a number of modifications were made. Several questions from Section II are considered inappropriate for the students, but for different reasons. Appendix H contains the first version of the test and its questions, whereas Appendix K presents the final version of the test. The questions the teacher considers inappropriate for the students in part or completely are questions 13, 15, 19, 43, 44, and 45 (see Appendix H). Question 13 requires students to describe a bicycle or a car, the student's apartment, house, or room. Describing a bicycle or car involves the use of very specific vocabulary such as spikes, gears, bumper, exhaust, starter, etc., which the students do not possess in German. Therefore, these parts were removed from the question. Question 15 requires the students to describe the city of Edmonton, however, considering the age of the students and the fact that they have a limited knowledge of landmarks, the question was changed: now it requires them to explain what they can show to a visitor from Germany, and what they could do together in Edmonton (see Appendix K, question 15). This change allows the students to include topics such as shopping, visiting a mall, going to the movies, going skating, and other activities. They do not have to know city landmarks. Based on S2's reaction and performance in the field testing, question 19 is removed entirely. S2 was asked to complain to a waiter in a restaurant about having been brought the wrong food. S2 had great difficulties formulating a complaint, and the teacher assumes that this type of situation is unrealistic for a child that age. It would be the child's parents who would launch a complaint, not the 12- or 13-year-old. The teacher considers question 43 also too difficult. She presumes that even though the students learn about the ancient Greeks, it is too big a step from knowing about the topic to imagining living during the times, and most importantly, expressing this hypothetical situation in German.

The most important change is the removal of Section III, the picture book narration. After listening to the rather halting and time-consuming narrations of all three students, we agreed on the inappropriateness of the task. The complexity of telling a story goes beyond the goal of testing the children's oral proficiency. Instead, it supplies information about children's storytelling abilities and skills, which make for fertile grounds for a linguistic study of that nature, but are not the focus of this oral proficiency test. Furthermore, the length of telling the complete story resulted in an overall test length of at least twenty minutes per student. This exceeds the desired test length of ten to fifteen minutes per student. As a result of the removal of the picture book, questions 44 and 45, which refer to the picture book directly, are removed from the cool-down section.

The removal of the picture book narration, which is the entire Section III, necessitated a replacement. We agreed that an additional task should be included in the test. We discussed the usefulness of describing a single picture instead, and agreed that it would demand from the students to use specific knowledge in German, such as expressing directions (e.g., prepositions such as 'next to', 'over', 'under', etc.), as well as using descriptive structures in the form of adjectives and adverbs (e.g., colours, size, descriptors, etc.). A single picture offers all of the above, and it would take the students only a few seconds to briefly describe one picture, instead of the four to six minutes necessary for the picture book narration. The form or source of the picture is not relevant and the teacher could, for example, use a photograph or a picture from a magazine.

So far, only two questions were added to the test. Appendix K shows that questions 35 and 38 have been added to subsection D, in which the students can choose what tense they use. Question 35 requires the students to compare how people live in Germany and Canada, and how their houses differ. This issue is discussed in class and requires topic-specific vocabulary as well as recalling factual information. This is one of the easier questions to be asked at the beginning or the middle of the school year and part of the cultural content questions. The more difficult question 38 requires students to compare life in China in the past and today. This topic is discussed in class, and students are expected to use topic-specific vocabulary in addition to the facts they learned.

A partial addition was made to question 41 (Appendix H), which becomes question 42 in the final test version presented in Appendix K: in addition to favourite rock star, rock group, and actor, we added a favourite sport celebrity to the list, because a lot of children at this age have a favourite basketball or hockey player.

This concludes the modifications to the first version of the test. The Grade Six teacher revealed that the German-English bilingual program is in the process of undergoing basic changes in the teaching of German culture. Curriculum changes are not completed yet, but it is to be expected that the changes will result in additional questions in which cultural content is tested. Clearly, questions can be added to any subsection of Sections I, II, and IV to address specific needs.

In summary, this chapter illustrated the processes necessary to design the oral test for the Grade Six students in the German-English bilingual school. The preliminary study offered insights into the language abilities of these children, and showed that certain criteria from tests presently in use fulfill the teacher's needs. The preliminary also exemplified, that two criteria the teacher wants included in the test, namely code-switching and cultural knowledge, need to be added to the test. Based on this preliminary study, the first version of the test was designed and used in a field testing. The field testing allowed for 'testing the test' and in the end, last modifications were made to the first test version, resulting in the final test.

CHAPTER V.

SUMMARY, DISCUSSION, AND IMPLICATIONS

In this final chapter of the thesis, I will summarize the expectations outlined in the first two chapters that were put on the test designed in this study, the results of the preliminary study, and the field testing of the first test version. The summary is followed by a discussion on the current oral proficiency test proposed here, its strengths and weaknesses, as well as its value to the teaching in the bilingual Grade Six. The possibility to adjust the test currently designed for Grade Six to the lower grades from kindergarten to Grade Five is briefly addressed. As a final point, this test has implications for bilingual schools at large, and gives rise to proposals for future FL oral proficiency tests.

5.0. SUMMARY

Chapter 2 outlined the terminology essential for the study, and the conditions necessary to design an oral proficiency test for the target group, namely Grade Six students in a German-English bilingual school. The goal of a proficiency test is that the test results illustrate how the foreign or second language is used for communication. Therefore, a situation must be created in which communication takes place. Furthermore, the proficiency test must reflect the learner's ability to handle real-life situations. Even though a situation presented in a test is never a real-life situation, it is the closest one can get to it. The questions in the final test version fulfill the goals of this proficiency test: the students have to use German to communicate feelings, factual knowledge, ideas, and opinions in a variety of situations.

The first research step stated in Chapter 1 determines the necessary requirements to design an evaluation method such as an oral proficiency test. In Chapter 2 the necessary requirements any test must fulfill are discussed. According to Finocchiaro and Sako (1983) and Bachman and Palmer (1996), a test has to be valid, reliable, comprehensible, and practical. Consequently, we must ask if the test designed in this study satisfies all of these requirements:

a.) The test designed for the Grade Six students fulfills content validity (the items correspond to course instructions) and construct validity (the test is constructed according to its goal). However, the concurrent validity has not been established, since the test has not been compared to an independent test. The test's predictive validity (the test results correspond with a student's previous performance) has also not been measured. The correspondence of the test results with a student's previous performance, is an issue left to the teacher to decide after multiple recordings. The test has to be used at the beginning of a school year in order to receive data to compare to the results from the year end test.

b.) The test is considered reliable, if it produces consistent results and gives consistent information: We have used the test only once with three students, and therefore the results can not be considered consistent. However, the test evaluations are consistent, and the overall final marks are consistent as well. Therefore, the test demonstrates inter-rater reliability.

c.) The test is considered comprehensive, if it contains an appropriate proportion of items from all aspects of the material to be tested: We can say for certain that the test contains items from all aspects: grammatical tenses and structures, a variety of topics and vocabulary, and cultural content.

d.) The test is considered practical, if it is low in cost, easy to score, and easy to administer: All of these requirements are met, since the only materials the teacher needs are paper, a tape-recorder, and a cassette, items which every school possesses. The test is easy to score because the teacher only has to match the SOPA criteria with the students' utterances on the tape. The test is easy to administer because the teacher can take one student at a time out of the classroom, while the rest of the class works on a project.

Consistent with Finocchiaro and Sako (1983), the test designed in this study meets the majority of these requirements. Bachman and Palmer (1996) added also authenticity, interactiveness, and impact to the issues above. Authenticity refers to the fact that the language used in the test must reflect the language used outside of the test. Interactiveness refers to the fact, that the individual learner's language abilities are reflected in the language used in the test, and impact refers to the impact the test has on society, the education system, and the individual. All three of these criteria are satisfied by the oral proficiency test designed in this study. The test's impact will be discussed later in this chapter.

The second and third steps named in Chapter 1 that needed to be taken enquire about oral proficiency tests presently in use, and their adequacy for the target group in the study here. Chapter 3 describes and discusses the advantages and disadvantages of existing oral tests, as well as the probability to apply these tests to the Grade Six students. Most of the existing tests are inadequate for this purpose, and the ETS scores were merely used for the preliminary study and proved otherwise insufficient. The SOPA criteria were found to include most of the categories that are of interest to our test, and the TOPT speaking tasks showed potential in order to elicit information. The main focus of the oral proficiency test designed in this study is to show what the students can do with the language (cf. Rhodes, 1996), and the SOPA criteria comply with this focal point for the most part. The TOPT questions were easily reworded to be suitable for children of that specific age, and the four parts used in the TOPT were also incorporated in the preliminary and the final version of the oral proficiency test in this study.

The preliminary study discussed in section 4.0. supplied the necessary information about the Grade Six children's German-speaking abilities in this bilingual elementary school. The observations revealed that within one grade level, students are at different oral proficiency levels, and that these levels can be evaluated and graded. Most importantly, the preliminary study exemplifies what these students are capable of in the foreign language, German. The necessary aspects of oral proficiency and communicative competency are accomplished at different proficiency levels. The three students of the preliminary study demonstrated to be fluent speakers of German, possess native-like pronunciation, and are able to talk about different topics using a variety of vocabulary, tenses, and grammatical structures. Moreover, the preliminary study confirmed that existing oral proficiency tests are not sufficient to describe these students' language abilities.

The results of the preliminary study and the field testing of the first test version revealed, that the oral proficiency tests presently in use are not adequate for the target group in this study. However, an entirely new test is not required: in the field testing the first test version was tried on three students of the current bilingual Grade Six and demonstrated, that the SOPA for young language learners offers the most useful evaluation criteria, and has to be supplemented with only two additional evaluation criteria, namely cultural content knowledge and amount and type of code-mixing. The SOPA already includes listening comprehension, fluency, vocabulary, and grammar criteria. The questions as compiled for Section II of the final test (Appendix K), allow the teacher to obtain information about the student's topic-related (content) and situation-related (context) abilities. The types of questions demand different skills from the students, for instance, describing, reporting, accounting facts, hypothesizing, or problem-solving. Furthermore, the task of telling a story in the third section of the first test version proved unproductive and ineffective, and is too time-consuming for the limited time available to teachers. The evaluation results and grades produced the desired outcome, namely the general agreement on the students' oral proficiency, but more importantly, the results support the test's validity and scoring reliability.

5.1. DISCUSSION

The oral proficiency test proposed in this thesis will be put into operation at the end of this 1999 – 2000 school year. The final test has its strengths and weaknesses. The advantages include:

- the majority of the essential criteria that must be met for a test to be reliable, comprehensive, and practical are established;

- the teacher's crucial input addresses the school's and the teacher's needs based on the curriculum key goals;
- the test is tape-recorded and the evaluation and grading can be done at a time convenient for the teacher;
- two teachers can share the work and save time: each teacher records half of the students in the class, and the other teacher evaluates them, and vice versa;
- the length necessary to test one student does not exceed 15 minutes;
- the test is flexible: questions can be modified (added, changed, or removed);
- the teacher can change the topic during the test if the data elicited is insufficient;
- the test is practical: economic (low in cost), easy to score, and easy to administer;
- the SOPA criteria include listening comprehension and fluency, as well as issues of accuracy such as vocabulary and grammar;
- the variety of questions allows for the testing of structures, tenses, cultural content and knowledge, factual information, and topic-specific vocabulary. Therefore, both linguistic and nonlinguistic abilities are included;
- accuracy is an important component, but not the focus of the test;
- the code-mixing criteria added to the test allows to observe possible language attrition processes of former monolingual German-speaking children who enter the program.

Clearly, the list compiling the test's advantages is extensive and seems rather convincing, however, the test exhibits some disadvantages:

- not all of the test requirements are proven yet: concurrent validity and predictive validity have not been tested. Therefore, the test's validity has not been established yet;
- if only one teacher uses the test, the recording time and the evaluation time add up to approximately 30 minutes per student;
- the warm-up and cool-down questions offer little variety yet;
- certain questions have not been tested in the preliminary test and may prove to be unsatisfactory. This may occur during an actual oral test application and may result in insufficient data at that time;
- the pictures or photographs used in Section III may prove to be unsuitable at times;
- according to the teacher, some students are very nervous when being tape-recorded while speaking German (test effect: anxiety). This may result in oral performances that are not accurate reflections of the student's true speaking abilities;
- due to the flexibility, the teacher may assist weaker students with easier questions which may influence the students' performances and result in an unjust representation of the student's abilities;

Not all of these disadvantages can be eliminated, and it is not my intention to try and design an infallible oral proficiency test. Advantages and disadvantages are present in every test, regardless of its careful construction. Nevertheless, the careful consideration of the observations and results of two test runs, the preliminary study and the field testing of the first test version, compose an oral proficiency test for the Grade Six students of this bilingual program that is useful, one that can be applied and scored with ease. The oral proficiency test and its results are a helpful tool to evaluate the overall language abilities of these children and is believed to be a most valuable element to establish proficiency in one of the four language skills, namely, speaking the foreign or second language (cf. Curtain & Pesola, 1994). Byrnes (1987) recognizes, that the language learner's proficiency level must be presented holistically, which means that the teachers must report all of the children's abilities to parents and children. This includes that the children's reading, writing, listening, and speaking abilities in German are evaluated and graded in order to establish the entire and all-inclusive proficiency level in German. I believe that the oral proficiency test designed in this study offers the appropriate data needed to help construct a holistic evaluation of these children's language abilities.

5.2. IMPLICATIONS

5.2.1. Implications for Bilingual Schools

The creation of a oral proficiency test based on an existing and long-established test that can be adjusted to the individual needs of a particular school is a desirable goal for foreign and second language school programs. As discussed in sections 2.1. and 2.2., a variety of language programs require different assessment devices and language tests. The trend observed in the testing methods of the last years stresses the importance of communicative abilities, which is a significant component of oral testing. It is in the interest of the bilingual school programs to promote and strengthen their programs, and with the help of an evaluation method such as a oral proficiency test, young learners' language abilities can be confirmed and substantiated in a reliable manner. One curriculum goal of many North American bilingual elementary programs is to teach a non-dominant language (cf. Curtain & Pesola, 1994, Lindholm & Fairchild, 1990, Robinson, 1978) to young language learners with the ultimate goal of becoming proficient speakers of this language. An oral test such as the one designed in this thesis enables teachers to look at students' utterances apart from the formal language and culture teaching during class, however, it does not eliminate form and cultural knowledge from the test. Instead, the children's oral skills are the focus of the evaluation and assessment, again, focusing on what these children are capable of doing with this foreign or second language, as well as illustrating how they convey their knowledge.

The Grade Six teacher of the thesis reported here observed during the evaluation that the grammar checklist, which serves as a reminder to grammatical features taught in class, must be considered differently during the evaluation of oral proficiency: a distinction must be made between linguistic knowledge and its application, and oral realization in the form of actual utterances. Even though the students receive formal lessons on the rules and structures of grammar early on and throughout the elementary school years, the bilingual program focuses on integrating German into daily routines and subject matters. As a result, it creates a more natural language environment and the students acquire German through what could be considered a partial immersion method. The main problem with the grammar checklist is that it illustrates all the grammar the students were taught, and are probably capable of recognizing and maybe even using in a grammar test. However, the spoken realization of this knowledge must account for something: spoken language and written language differ greatly, and we are dealing with the spoken language in an oral proficiency test. Immediate communication and transfer of a message are the focus points in a conversation, or in answering questions. Tenses, structures, vocabulary, cultural knowledge, fluency, and pronunciation are mere facilitators to relay a message, to communicate. Therefore, it is pertinent that schools offering a bilingual program realize this difference, and evaluate the children accordingly, which, I believe, is accomplished by using the SOPA.

It is possible to make changes to the existing oral proficiency test, its evaluation criteria and grading, in order to use variations of the test with the children of the lower grade levels, from kindergarten to Grade Five. To include these changes goes beyond the scope of the study reported here, however, the teachers of these grade levels are certainly able to make all necessary changes themselves, and I am available to help them. In accordance with this, any German-English bilingual school is able to use the test with possible variations for their Grade Six.

5.2.2. Implications for Future FL-Test

Many oral tests have been designed and applied to a variety of language learners, and I mentioned only a handful of the tests and grading scales presently in use. The target groups for these tests always differ. Children of all ages are tested for their language abilities, from elementary school to high school. The adult learners involve even more variety: the FL-and L2-learners include university students, college students, people who desire to learn another language, immigrants, and many more. The different learners have different motivations for learning a foreign or second language, and all of them have different goals. Based on the learner target group and the institution applying the test, modifications can be made to existing tests. The oral proficiency test designed in this study is created for one particular target group: Grade Six students of a German-English bilingual school in Alberta, Canada. The test is based on the SOPA and the TOPT, but it is, nevertheless, a new test, as it contains two new elements (code-mixing and cultural knowledge) not present in other tests, for example the OPI. Any oral proficiency test for FL-learners can be designed following this model. The goals and objectives of what needs to be tested must be clear, and the test results must measure what the test sets out to measure. Whereas no new revelations have taken place during this study, one important observation was made nonetheless: the amount and type of code-mixing that a particular group may use should be taken into account, if it is revealed to be a significant feature, as with the Grade Six students reported on in this thesis. I believe it to be an essential measure that helps evaluate foreign or second language abilities. Further research in this area should investigate a possible correlation between code-mixing and oral proficiency.

References

Alberta Education (1996). *Curriculum Handbook for Parents: Grade 4 (Edmonton Public Schools.* Alberta: Alberta Education.

American Council on the Teaching of Foreign Languages (1999). *ACTFL Proficiency Guidelines – Speaking*. New York: ACTFL Inc.

Apeltauer, E. (1993). *Zweit- und Fremdsprachenerwerbstheorien* (vorläufige Arbeitsversion). Gesamthochschule Kassel (GhK).

Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. & Palmer, A. (1984). Some comments on the terminology of language testing. In Charlene Riviera (Ed.), *Communicative Competence Approaches to Language Proficiency Assessment: Research and Application*. 34- 43. Clevedon: Multilingual Matters.

Bachman, L. & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Baetens Beardsmore, H. (1986). *Bilingualism: Basic Principles (second edition)*. Clevedon: Multilingual Matters.

Bamberg, M. (1997). *The Acquisition of Narratives*. Berlin: Mouton de Gruyter.

Barnwell, D. (1993). Oral proficiency testing and the bilingual speaker. In Ana Roca and John M. Lipski (Eds.), *Spanish in the United States*, 199-209. Berlin: Mouton de Gruyter.

Berman, R. & Slobin, D. (1994). *Relating Events in Narrative: a Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Buck, K. , Byrnes, H., & Thompson, L., Eds. (1989). *ACTFL Oral Proficiency Interview Tester Training Manual*. New York.

Byrnes, H. (1985). *Second Language Acquisition: Insights from a Proficiency Orientation.* Paper presented at the American Council on the Teaching of Foreign Languages Annual Meeting, New York, November 1985.

Byrnes, H. (1987). Proficiency as a framework for research in second language acquisition. *The Modern Language Journal*, *71*, i, 44-49.

CAL Center for Applied Linguistics (1999). *About CAL*. http://www.cal.org/public/about.htm.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*, 1, 1-47.

Clyne, M. (1985). Medium or object – different contexts of (school-based) second language acquisition. In Kenneth Hyltenstam and Manfred Pienemann (Eds.). *Modelling and Assessing Second Language Acquisition*. Clevedon: Multilingual Matters.

College Board (1993). *The 1992 Advanced Placement Examination in German Language and Its Grading.* New Jersey: College Entrance Examination Board.

Cummins, J. (1984). Linguistic minorities and multicultural policy in Canada. In J. Edwards (Ed.). *Linguistic Minorities, Policies, and Pluralism*. London: Academic Press.

Curtain, H. & Pesola, C. (1994). *Languages and Children: Making the Match*. New York: Longman.

Dicks, J. & Rehorick, S. (1995). Evaluation: oral communication assessment and the portfolio concept. In Georges Duquette (Ed.), *Second Language Practice: Classroom Strategies for Developing Communicative Competence*, 134-148. Clevedon: Multilingual Matters.

Edmonton Public Schools (1991). *Outcomes, Expectations, and Indicators: Division II.* Edmonton: Edmonton Public Schools.

Edmonton Public School Board (1995). *Board Policies and Regulations: HK.AR Communicating Student Achievement and Growth*. Edmonton: Edmonton Public Schools.

Ellis, R. (1985). *Understanding Language Acquisition*. Oxford: Oxford University Press.

Fillmore, C. (1979). On fluency. In Charles J. Fillmore, Daniel Kempler and William S.-Y. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior*. New York: Academic Press.

Finocchiaro, M. & Sako, S. (1983). *Foreign Language Testing: A Practical Approach.* New York: Regents.

Galloway, V. (1980). Perceptions of the communicative efforts of American students of Spanish. *The Modern Language Journal*, *64*, 428-433.

Genesee, F. (1987). *Learning Through Two Languages: Studies of Immersion and Bilingual Education*. Cambridge: Newbury House.

Gutstein, S. & Goodwin, S. (1987). *The CLEAR Oral Proficiency Exam (COPE) Project Report[and] Addendum: Clinical Testing and Validity and Dimensionality Studies.* Michigan State University: Center for Language Education and Research.

Heilenman, L. & Kaplan, I. (1985). Proficiency in practice: the foreign language curriculum. In Charles J. James (Ed.), *Foreign Language Proficiency in the Classroom and Beyond*, 55-78. ACTFL Foreign Language Education Series, Skolje, IL: National Textbook.

Kuo, J. & Jiang, X. (1997). Assessing the assessments: the OPI and the SOPI. *Foreign Language Annals, 30*, 4, 503-512.

Lindholm, K. (1990). Bilingual immersion education: criteria for program development. In Amado M. Padilla, Halford H. Fairchild and Concepcion M. Valadez (Eds.). *Bilingual Education: Issues and Strategies*, 91-105. Newbury Park: Sage.

Lindholm, K. & Fairchild, H. (1990). Evaluation of an elementary school bilingual immersion program. In Amado M. Padilla, Halford H. Fairchild and Concepcion M. Valadez (Eds.). *Bilingual Education: Issues and Strategies,* 126-136. Newbury Park: Sage.

Loughrin-Sacco, S. (1990). The reflections and confessions of an ACTFL Oral Proficiency Interview workshop participant. *The Canadian Modern Language Review, 46*, 4, 706-713.

Malakoff, M. & Hakuta, K. (1990). History of language minority education in the United States. In Amado M. Padilla, Halford H. Fairchild and Concepcion M. Valadez (Eds.). *Bilingual Education: Issues and Strategies*, 27-43. Newbury Park: Sage.

Mayer, M. (1977). *Frog goes to Dinner*. Hong Kong: Puffin Pied Piper.

Padilla, A. (1994). *Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM)*. Available from Dr. Amado Padilla, Stanford University, School of Education, CA)

Perron, G. (1968). Testing spoken language: some unsolved problems. In A. Davies (Ed.). *Language Testing Symposium*, 107-116. London: Oxford University Press.

Politzer, R. (1978). Errors of English speakers of Germans as perceived and evaluated by German natives. *The Modern Languages Journal*, *62*, 253-261.

Robinson, N. (1978). *Language Management in Education: The Australian Context*. London: Allen and Unwin.

Rhodes, N. (1996). *Alternative Assessment for Immersion Students: The Student Oral Proficiency Assessment (SOPA)*. Paper presented at the European Conference on Immersion Programmes, Barcelona, Spain, September 25-28, 1996.

Rhodes, N. & Thompson, L. (1990). An oral assessment instrument for immersion students: COPE. In Amado M. Padilla, Halford H. Fairchild and Concepcion M. Valadez (Eds.), *Foreign Language Education: Issues and Strategies*, 75-94. Newbury Park: Sage.

Savignon, S. (1985). Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *The Modern Language Journal*, *69*, 2, 129-134.

Sieloff, S. (1985). From achievement toward proficiency through multi-sequence evaluation. In Charles J. James (Ed.), *Foreign Language Proficiency in the Classroom and Beyond*, 117-145. The ACTFL Foreign Language Education Series, Lincolnwood, IL: National Textbook Company.

Spolsky, B. (1978). Language testing: art or science. In Gerhard Nickel (Ed.), *Language Testing*, 215-234. Stuttgart: Hochschulverlag.

Stansfield, C. & Kenyon, D. (1991). *Development of the Texas Oral Proficiency Test (TOPT). Final Report*. Center for Applied Linguistics, Washington, DC.

Swender, E. & Duncan, G. (1998). ACTFL performance guidelines for K-12 learners. *Foreign Language Annals*, *31*, no. 4, 479-491.

Thompson, L. (1996). *CAL Student Self Assessment (SSA)*. (Available from Lynn Thompson, Center for Applied Linguistics, Washington, DC)

Thompson, L., Christian, D., Stansfield, C., & Rhodes, N. (1990). Foreign language instruction in the United States. In Amado M. Padilla, Halford H. Fairchild and Concepcion M. Valadez (Eds.), *Foreign Language Education: Issues and Strategies*, 22-35. Newbury Park: Sage.

Thompson, L., Boyson, B., & Rhodes, N. (2000). *Student Oral Proficiency Assessment (SOPA): A Foreign Language Listening and Speaking Assessment for Children Grades K-7. Administrator's Manual (Draft).* Washington: CAL.

Thompson, L., Kenyon, D., & Rhodes, N. (2000). *A Validation Study of the Student Oral Proficiency Assessment (SOPA)*. Unpublished draft from Center for Applied Linguistics, Washington for Iowa State University.

Van Hest, E., Poulisse, N., & Bongaerts, T. (1997). Self-repair in L1 and L2 production: an overview. *ITL*, *117-118*, 85-115.

APPENDIX A

The ACTFL Proficiency Guidelines – Speaking

(1986; revised 1999)

SUPERIOR

Speakers at the Superior level are able to communicate in the language with accuracy and fluency in order to participate fully and effectively in conversations on a variety of topics in formal and informal settings from both concrete and abstract perspectives. They discuss their interests and special fields of competence, explain complex matters in detail, and provide lengthy and coherent narrations, all with ease, fluency, and accuracy. They explain their opinions on a number of topics of importance to them, such as social and political issues, and provide structured argument to support their opinions. They are able to construct and develop hypotheses to explore alternative possibilities. When appropriate, they use extended discourse without unnaturally lengthy hesitation to make their point, even when engaged in abstract elaborations. Such discourse, while coherent, may still be influenced by the Superior speakers' own language patterns, rather than those of the target language.

Superior speakers command a variety of interactive and discourse strategies, such as turn-taking and separating main ideas form supporting information through the use of syntactic and lexical devices, as well as intonational features such as pitch, stress and tone. They demonstrate virtually no pattern of error in the use of basic structures. However, they may make sporadic errors, particularly in low-frequency structures and in some complex high-frequency structures more common to formal speech and writing. Such errors, if they do occur, do not distract the native interlocutor or interfere with communication.

ADVANCED HIGH

Speakers at the Advanced-High level perform all Advanced-level tasks with linguistic ease, confidence and competence. They are able to consistently explain in detail and narrate fully and accurately in all time frames. In addition, Advanced-High speakers handle the tasks pertaining to the Superior level but cannot sustain performance at that level across a variety of topics. They can provide a structured argument to support their opinions, and they may construct hypotheses, but patterns of error appear. They can discuss some topics abstractly, especially those relating to their particular interests and special fields of expertise, but in general, they are more comfortable discussing a variety of topics concretely.

Advanced-High speakers may demonstrate a well-developed ability to compensate for an imperfect grasp of some forms or for limitations in vocabulary by the confident use of communicative strategies, such as paraphrasing, circumlocution, and illustration. They use precise vocabulary and intonation to express meaning and often show great fluency and ease of speech. However, when called on to perform the complex tasks associated with the Superior level over a variety of topics, their language will at times break down or prove inadequate, or they may avoid the tasks altogether, for example, by resorting to simplification through the use of description or narration in place of argument or hypothesis.

ADVANCED MID

Speakers at the Advanced-Mid level are able to handle with ease and confidence a large number of communication tasks. They participate active un most informal and some formal exchanges on a variety of concrete topics relating to work, school, home, and leisure activities, as well as to events of current, public, and personal interest or individual relevance.

Advanced-Mid speakers demonstrate the ability to narrate and describe in all major time frames (past, present, and future) by providing a full account, with good control of aspect, as they adapt flexibly to the demands of the conversation. Narration and description tend to be combined and interwoven to relate relevant and supporting facts in connected, paragraph-length discourse.

Advanced-Mid speakers can handle successfully and with relative ease the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine situation or communicative task with which they are otherwise familiar. Communicative strategies such as circumlocution or rephrasing are often employed for this purpose. The speech of Advanced-Mid speakers performing Advanced-level tasks is marked by substantial flow. Their vocabulary is fairly extensive although primarily generic in nature, except in the case of a particular area of specialization or interest. Dominant language discourse structures tend to recede, although discourse may still reflect the oral paragraph structure of their own language rather than that of the target language.

Advanced-Mid speakers contribute to conversations on a variety of familiar topics, dealt with concretely, with much accuracy, clarity and precision, and they convey their intended message without misrepresentation or confusion. They are readily understood by native speakers unaccustomed to dealing with non-natives. When called on to perform functions or handle topics associated with the Superior level, the quality and/or quantity of speech will generally decline. Advanced-Mid speakers are often able to state an opinion or cite conditions; however, they lack the ability to consistently provide a structured argument in extended discourse. Advanced-Mid speakers may use a number of delaying strategies, resort to narration, description, explanation or anecdote, or simply attempt to avoid the linguistic demands of Superior-level tasks.

ADVANCED LOW

Speakers at the Advanced-Low level are able to handle a variety of communicative tasks, although somewhat haltingly at times. They participate actively in most informal and limited number of formal conversations on activities related to school, home, and leisure activities and, to a lesser degree, those related to events of work, current, public, and personal interest or individual relevance.

Advanced-Low speakers demonstrate the ability to narrate and describe in all major time frames (past, present and future) in paragraph length discourse, but control of aspect may be lacking at times. They can handle appropriately the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine situation or communicative task with which they are otherwise familiar, through at times their discourse may be minimal for the level and strained. Communicative strategies such as rephrasing and circumlocution may be employed in such instances. In their narrations and descriptions, they combine and link sentences into connected discourse of paragraph length. When pressed for a fuller account, they tend to grope and rely on minimal discourse. Their utterances are typically not longer than a single paragraph. Structure of the dominant language is still evident in the use of false cognates, literal translations, or the oral paragraph structure of the speaker's own language rather than that of the target language.

While the language of Advanced-Low speakers may be marked by substantial, albeit irregular flow, it is typically somewhat strained and tentative, with noticeable self-correction and a certain 'grammatical roughness.' The vocabulary of Advanced-Low speakers is primarily generic in nature.

Advanced-Low speakers contribute to the conversation with sufficient accuracy, clarity, and precision to convey their intended message without misrepresentation or confusion, and it can be understood by native speakers unaccustomed to dealing with non-natives, even though this may be achieved through repetition and restatement. When attempting to perform functions or handle topics associated with the Superior level, the linguistic quality and quantity of their speech will deteriorate significantly.

INTERMEDIATE HIGH

Intermediate-High speakers are able to converse with ease and confidence when dealing with most routine tasks and social situations of the Intermediate level. They are able to handle successfully many uncomplicated tasks and social situations though hesitation and errors may be evident.

Intermediate-High speakers handle the tasks pertaining to the Advanced level, but they are unable to sustain performance at that level over a variety of topics. With some consistency, speakers at the Intermediate High level narrate and describe in major time frames using connected discourse of paragraph length. However, their performance of these Advanced-level tasks will exhibit one or more features of breakdown, such as the failure to maintain the narration or description semantically or syntactically in the appropriate major time frame, the disintegration of connected discourse, the misuse of cohesive devises, a reduction in breadth and appropriateness of vocabulary, the failure to successfully circumlocute, or a significant amount of hesitation.

Intermediate-High speakers can generally be understood by native speakers unaccustomed to dealing with non-natives, although the dominant language is still evident (e.g. use of code-switching, false cognates, literal translations, etc.), and gaps in communication may occur.

INTERMEDIATE MID

Speakers at the Intermediate-Mid level are able to handle successfully a variety of uncomplicated communicative tasks in straightforward social situations. Conversation is generally limited to those predictable and concrete exchanges necessary for survival in the target culture; these include personal information covering self, family, home, daily activities, interests and personal preferences, as well as physical and social needs, such as food, shopping, travel and lodging.

Intermediate-Mid speakers tend to function reactively, for example, by responding to direct questions or requests for information. However, they are capable of asking a variety of questions when necessary to obtain simple information to satisfy basic needs, such as directions, prices and services. When called on to perform functions or handle topics at the Advanced level, they provide some information but have difficulty linking ideas, manipulating time and aspect, and using communicative strategies, such as circumlocution.

Intermediate-Mid speakers are able to express personal meaning by creating with the language, in part by combining and recombining known elements and conversational input to make utterances of sentence length and some strings of sentences. Their speech may contain pauses, reformulations and self-corrections as they search for adequate vocabulary and appropriate language forms to express themselves. Because of inaccuracies in their vocabulary and/or pronunciation and/or grammar and/or syntax, misunderstandings can occur, but Intermediate-Mid speakers are generally understood by sympathetic interlocutors accustomed to dealing with non-natives.

INTERMEDIATE LOW

Speakers at the Intermediate-Low level are able to handle successfully a limited number of uncomplicated communicative tasks by creating with the language in straightforward social situations. Conversation is restricted to some of the concrete exchanges and predictable topics necessary for survival in the target language culture. These topics relate to basic personal information covering, for example, self and family, some daily activities and personal preferences, as well as to some immediate needs, such as ordering food and making simple purchases. At the Intermediate-Low level, speakers are primarily reactive and struggle to answer direct questions or requests for information, but they are also able to ask a few appropriate questions.

Intermediate-Low speakers express personal meaning by combining and recombining into short statements what they know and what they hear from their interlocutors. Their utterances are often filled with hesitancy and inaccuracies as they search for appropriate linguistic forms and vocabulary while attempting to give form to message. Their speech is characterized by frequent pauses, ineffective reformulations and self-corrections. Their pronunciation, vocabulary and syntax are strongly influenced by their first language but, in spite of frequent misunderstandings that require repetition or rephrasing, Intermediate-Low speakers can generally be understood by sympathetic interlocutors, particularly those accustomed to dealing with non-natives.

NOVICE HIGH

Speakers at the Novice-High level are able to handle a variety of tasks pertaining to the Intermediate level, but are unable to sustain performance at that level. They are able to manage successfully a number of uncomplicated communicative tasks in straightforward social situations. Conversation is restricted to a few of the predictable topics necessary for survival in the target language culture, such as basic personal information, basic objects and a limited number of activities, preferences and immediate needs. Novice-High speakers respond to simple, direct questions or requests for information; they are able to ask only a very few formulaic questions when asked to do so.

Novice-High speakers are able to express personal meaning by relying heavily on learned phrases or recombinations of these and what they hear from their interlocutor. Their utterances, which consist mostly of short and sometimes incomplete sentences in the present, may be hesitant or inaccurate. On the other hand, since these utterances are frequently only expansions of learned material and stock phrases, they may sometimes appear surprisingly fluent and accurate. These speakers' first language may strongly influence their pronunciation, as well as their vocabulary and syntax when they attempt to personalize their utterances. Frequent misunderstanding may arise but, with repetition or rephrasing, Novice-High speakers can generally be understood by sympathetic interlocutors used to non-natives. When called on to handle simply a variety of topics and perform functions pertaining to the Intermediate level, a Novice-High speaker can sometimes respond in intelligible sentences, but will not be able to sustain sentence level discourse.

NOVICE MID

Speakers at the Novice-Mid level communicate minimally and with difficulty by using a number of isolated words and memorized phrases limited by the particular context in which the language has been learned. When responding to direct questions, they may utter only two or three words at a time or an occasional stock answer. They pause frequently as they search for simple vocabulary or attempt to recycle their own and their interlocutor's words. Because of hesitations, lack of vocabulary, inaccuracy, or failure to respond appropriately, Novice-Mid speakers may be understood with great difficulty even by sympathetic interlocutors accustomed to dealing with non-natives. When called on to handle topics by performing functions associated with the Intermediate level, they frequently resort to repetition, words from their native language, or silence.

NOVICE LOW

Speakers at the Novice-Low level have no real functional ability and, because of their pronunciation, they may be unintelligible. Given adequate time and familiar cues, they may be able to exchange greetings, give their identity, and name a number of familiar objects from their immediate environment. They are unable to perform functions or handle topics pertaining to the Intermediate level, and cannot therefore participate in a true conversational exchange.

Reprinted by permission of ACTFL, the copyright owners.

APPENDIX B

Nine-level Rating Scale for Student Oral Proficiency Assessment (SOPA)

Adapted from the COPE

	JR NOVICE LOW	JR NOVICE MID	JR NOVICE HIGH
Listening Comprehension	Recognizes isolated words and high frequency expressions.	Understands predictable questions, statements, and commands in familiar topic areas (with strong contextual support), though at slower than normal rate of speech and/or with repetitions.	Understands simple questions, statements and commands in familiar topic areas, and some new sentences with strong contextual support. May require repetition, slower speech, or rephrasing.
Fluency	Produces only isolated words in very specific topic areas and/or high-frequency expressions such as *good morning*. Has essentially no functional communicative ability.	Uses a limited number of words and phrases within predictable topic areas. May also use longer memorized expressions.	Uses high frequency expressions and other memorized expressions with reasonable ease. Signs of originality and spontaneity begin to emerge. Attempts to create sentences with some success, but is unable to sustain sentence level speech.
Vocabulary	Uses words in very specific topic areas in predictable contexts and a few memorized, high frequency expressions.	Uses vocabulary consisting of specific words, high-frequency expressions, and/or other longer, memorized expressions in a limited number of topic areas. Frequent searches for words are common.	Uses vocabulary centering on basic objects, places, and common kinship terms, adequate for minimally elaborating utterances in predictable topic areas.
Grammar	May use memorized, high-frequency phrases accurately, but lacks awareness of grammar and syntax.	May use memorized expressions accurately. Attempts at putting together two or three word phrases may also be accurate. Does not use verbs to speak creatively in sentence-level speech.	Relies mainly on memorized expressions, and may use them accurately. Creates some sentences-level speech with conjugated verbs, but in general, verbs are lacking or are not conjugated.

	JR INTERMEDIATE LOW	JR INTERMEDIATE MID	JR INTERMEDIATE HIGH
Listening Compre-hension	Understands new sentence-level questions and commands in a limited number of content areas with strong contextual support. Follows conversation at a fairly normal rate of speech.	Understands sentence-level utterances in new contexts though some slow-downs may be necessary. Carries out commands without prompting. May show some difficulty on unfamiliar topics.	Usually understands speech at normal speed. Seldom has comprehension problems on everyday topics. Understands longer stretches of connected discourse on a number of topics.
Fluency	Able to handle a limited number of everyday social and academic interactions. Maintains simple conversations at the sentence level by creating with the language, although in a restrictive and reactive manner.	Shows evidence of spontaneity in conversation. May initiate talk without relying on questions or prompts. Gives simple descriptions successfully. May attempt longer, increasingly complex sentences. Few, if any connectors are used.	Maintains conversation with increasing fluency. Uses language creatively to initiate and sustain talk. There is emerging evidence of connected discourse in descriptions and simple narratives.
Vocabulary	Has sufficient vocabulary for making statements and asking questions to satisfy basic social and academic needs, but not for explaining or elaborating on them.	Has basic vocabulary permitting discussion of a personal nature and limited academic topics. Serious gaps may exist for speaking about topics of general interest. Attempts at circumlocution may be ineffective.	Has a broad enough vocabulary for discussing simple social and academic topics in generalities, but lacks detail. Sometimes achieves successful circumlocution.
Grammar	Goes beyond memorized utterances to create with the language at the sentence level. Verbs are conjugated, but may be inaccurate. Speech may also contain many other grammatical inaccuracies.	Maintains simple conversation, mostly in present tense, although an awareness of other tenses may be evident. Many grammatical inaccuracies are present.	Uses present tense, but lacks control of past tenses. Grammatical inaccuracies are present.

	JR ADVANCED	JR ADVANCED-HIGH	SUPERIOR
Listening Comprehension	Understands main ideas and most details in connected speech on a variety of topics, and is aware of connectors, but may be unable to follow complicated oral speech. May have difficulty with highly idiomatic speech.	Understands complex academic speech and highly idiomatic conversation, though confusion may occur due to sociocultural nuances or unfamiliar topics.	Has no difficulty understanding social or academic interactions. Follows extended speech in lectures, speeches, and reports, having an awareness of underlying organizational patterns.
Fluency	Reports facts easily. Can discuss topics of personal interest and some topics, satisfying the requirements of school and every day situations. Narrates and describes successfully. Connects sentences smoothly, and organizes speech using sequencing devices such as the first, next, finally, etc.	Handles most academic and social requirements with confidence, but may demonstrate inadequacy under the demands of complex tasks. Emerging ability to support opinions and hypothesize on abstract topics is evident.	Uses a variety of interactive strategies to participate effectively in formal and informal, social and academic situations. Can support opinions, hypothesize on abstract concepts, and adjust speech to the type of audience.
Vocabulary	Has vocabulary adequate for including some detail when talking about concrete or factual topics of a personal nature and of general interest. May use a variety of idiomatic expressions. Can use circumlocution effectively.	Has sufficient vocabulary to discuss a wide variety of topics related to everyday social and academic situations. Lack of vocabulary rarely interrupts the flow of speech.	Is familiar with idiomatic expressions and less common vocabulary which permit discussing topics in unfamiliar situations. Vocabulary is extensive.
Grammar	Has good control in expressing present, past, and future times. Errors, if present, do not interfere with communication. Non-standard varieties of grammar may be used (native speakers).	Uses conditionals, but can not sustain a hypothetical mode with appropriate verb forms. Uses most grammatical structures accurately, but patterns of error may persists.	Uses complex forms as well. Occasional errors in complex, low-frequency structures may occur, but they do not interfere with communication. There are no major patterns of error.

Reprinted by permission of the Center for Applied Linguistics, the copyright owners.

APPENDIX C

TOPT Speaking Tasks

1. Introduce Yourself
   Be able to give your name and basic personal information such as would be given at a first meeting.

2. Explain a Familiar, Simple Process
   Be able to explain how to accomplish everyday processes such as writing a check, borrowing a book form the library, or taking attendance in the classroom.

3. Describe a Sequence of Events in the Past
   Be able to use and sequence language indicating past time in order to narrate an event or incident which occurred recently.

4. Propose and Defend a Course of Action with Persuasion

   In light of at least two possible choices of action, be able to propose and defend a course of action in such a way as to persuade others to accept your choice.

5. Describe Typical Routines
   Be able to use and sequence language indicating present or habitual time in order to narrate recurring events or routines, everyday activities, etc.

6. Make Purchases
   Be able to request items, discuss prices, and handle currency in a situation involving a purchase.

7. Talk About Personal Activities
   Be able to talk about your leisure activities, favorite pastimes, and preferred hobbies.

8. Hypothesize About an Impersonal Topic
   Be able to discuss various possibilities ("what if" situations) surrounding an abstract, impersonal topic.

9. Talk About Family Members
   Be able to give the names of the members of your family and simple descriptive information, such as their occupations and physical characteristics.

10. Give a Brief, Organized Factual Summary
    Be able to summarize in an "oral report" fashion factual information about topics of a personal or professional nature.

11. State Your Personal Point of View on a Controversial Subject
Be able to state what you believe on a controversial subject and why you hold those beliefs.

12. Describe Expected Future Events
Be able to use and sequence language indicating future time in order to narrate expected occurrences of a personal nature, such as a planned trip or activity.

13. Explain a Complex Process in Detail
Be able to explain in detail a non-routine process of an impersonal nature, such as how to carry out a scientific investigation or how to write a term paper.

14. Order a Meal
Be able to ask questions about menu items, order food, and ask for and settle a bill.

15. Express Personal Apologies
Be able to apologize clearly and appropriately to an offended party.

16. Give Advice
Be able to give advice to someone faced with making a decision between two or more choices, giving supporting reasons for the advice given.

17. Hypothesize About a Personal Situation
Be able to say what you would do in a hypothetical situation.

18. Describe Your Daily Routine
Be able to narrate your typical daily activities.

19. Give Instructions
Be able to give instructions and explain the steps involved in carrying out an activity.

20. Give a Brief Personal History
Be able to talk about your personal background.

21. State Advantages and Disadvantages
Be able to state the advantages and disadvantages of a situation (such as living in a big city), a decision (such as going to college), or an object that has affected society (such as the computer).

22. Support Opinions
Be able to state, support and defend a personally-held opinion or belief about an issue.

23. Describe Health Problems
Be able to describe health problems or conditions.

24. Discuss a Professional Topic
Be able to discuss at length and in detail a topic of professional interest.

25. Describe a Complex Object in Detail
Be able to describe a complex object as a car or bicycle in detail and with precise vocabulary.

26. Lodge a Complaint
Be able to lodge a complaint, giving the reasons for and details behind the complaint.

27. Talk About Your Future Plans
Be able to state and describe your personal or professional plans, goals and ambitions.

28. Give a Professional Talk
Be able to present a talk on a topic of professional interest.

29. Make Arrangements for Future Activities
Be able to inquire about and to make arrangements for future activities, and to set the date, time and place.

30. Evaluate Issues Surrounding a Conflict
Be able to present arguments on both sides of a familiar issue or topic and evaluate their relative merits.

31. Give Directions
Be able to give directions on how to get from one place to another.

32. Describe a Place
Be able to describe in detail a particular place, such as a school, a store, or a park.

33. Explain a Complex Process of Personal Nature
Be able to describe and explain in detail a non-routine process such as how to get a job, or how to apply to college.

34. Hypothesize About Probable Outcomes
Be able to discuss what could happen if something unexpected occurs.

35. Correct an Unexpected Situation
Be able to handle an unexpected outcome, such as receiving faulty merchandise.

36. Change Someone's Behavior through Persuasion
Be able to persuade someone to do something he or she is not inclined to do, or to cease doing something which is annoying to you.

37. Describe Habitual Actions in the Past
Be able to describe people, places or things in the past, such as the work schedule you used to have or leisure activities you used to do.

38. Compare and Contrast Two Objects or Places
Be able to compare and contrast two objects, places, or customs.

Reprinted by permission of the Center for Applied Linguistics, the copyright owners.

The material on page 106 has been removed due to copyright restrictions.

The information removed was Appendix D: ETS Scoring Criteria, and was taken from the following original source:
College Board (1993). *The 1992 Advanced Placement Examination in German Language And Its Grading.* New Jersey: College Entrance Examination Board. Pages 62-64.

The ETS Scoring Criteria contain descriptions on scoring directed responses and picture sequences.

The material on page 107 has been removed due to copyright restrictions.

The information removed was Appendix D: ETS Scoring Criteria, and was taken from the following original source:
College Board (1993). *The 1992 Advanced Placement Examination in German Language And Its Grading.* New Jersey: College Entrance Examination Board. Pages 62-64.

The ETS Scoring Criteria contain descriptions on scoring directed responses and picture sequences.

APPENDIX E

MOCAP Evaluation Sheet and Global Descriptors of Levels of Performance

Part A: Evaluation Sheet

I. Communication of message

A = Message communicated effectively and appropriately.
B = Message communicated but effectiveness and/or accuracy must be improved.
C = Message not communicated effectively.

Language tasks:	A	B	C
1. *Give information about location*	☐	☐	☐
2. *Give numerical information*	☐	☐	☐
3. *Give information about an event in the past*	☐	☐	☐

II. DIAGNOSTIC ANALYSIS

Articulation: pronunciation,; intonation; fluency ____________________
Facility: vocabulary (precision and variety); communicative strategies ____
Grammaire: grammatical precision; discourse cohesion _______________

III. GLOBAL EVALUATION A ☐ B ☐ C ☐
(see global descriptors in Part B below)

Section I. With respect to language tasks, each task to be accomplished in a given evaluation technique is listed on the evaluation scheme which accompanies that technique. Teachers are asked to identify whether this task has been accomplished completely (A), in part (B), or not al all (C).

Section II. The scheme also allows for the measure of student progress and a diagnostic assessment of strengths and weaknesses. A section is provided for analysis of students' linguistic, social-linguistic, discourse, and strategic competence and space is provided for the writing of specific comments with respect to student performance in these areas. In this section, teachers are requested to identify major strengths and weaknesses. Since one cannot evaluate everything nor can one diagnose each and every strength or weakness, the aim is to provide students with valuable feedback with respect to their progress (positive feedback), and one or two areas where they need to improve (negative feedback).

Section III. The evaluation scheme also allows for a global, holistic evaluation of student performance. This is arrived at by considering the evaluation with respect to the language tasks and individual competencies discussed above, and by comparing student performance to the descriptions provided of differing levels of competence – A, B, or C. These descriptors are located [...] [in Part B below].

Part B: Global descriptors of levels of performance

Level A. The learner is almost always capable of using the language to carry out the target language tasks and does so in an appropriate manner. He/she almost always expresses him/herself with a superior level of grammatical precision pronunciation, intonation and flow never interfere with communication. The learner almost always understands without the need for repetition or reformulation. Communicative strategies such as gestures, requests for repetition, circumlocutions are almost always understandable, complete and naturally integrated into conversation. Discourse is logical and coherent. Discourse connectors such as pronouns and transitional words are used correctly and systematically.

Level B. The learner is often capable of using the language to carry out the target language tasks and often does so in an appropriate manner. He/she often expresses him/herself with an average level of grammatical precision. In general, pronunciation, intonation and flow never interfere with communication. The learner often understandable without the need for repetition or reformulation. Communicative strategies such as gesture, requests for repetition, circumlocutions are often understandable, complete naturally integrated into conversation. Discourse is often logical and coherent. Discourse connectors such as pronouns and transitional words are often used correctly and systematically.

Level C. The learner is often incapable of using the language to carry out the target language tasks. He/she expresses him/herself with little grammatical precision. Pronunciation, intonation and flow often interfere with communication. The learner does not understand without repetition or reformulation. Communicative strategies such as gestures, requests for repetition, circumlocutions are rarely used. Discourse is often illogical and incoherent. Discourse connectors such as pronouns and transitional words are rarely used correctly.

[source: Dicks & Rehorick 1995]

APPENDIX F

Preliminary Study: Instructions and Questions

(All of the questions were used as guidelines. The actual wording changed every time)

Part I:

Kannst du mir ein bisschen von dir und deiner Familie erzählen? Wie viele Geschwister hast du, was machst du gern, hast du ein Hobby?

('Can you tell me a little bit about yourself and your family? How many brothers and sisters do you have, what do you like to do, do you have any hobbies?')

Part II:

(not all of the questions were asked, but every student answered a) and b))

a) Stell dir vor, du würdest eine Millionen Dollar gewinnen. Was würdest du mit dem Geld machen?

('Imagine you were to win one million dollars. What would you do with the money?')

b) Wohin möchtest du am liebsten reisen und warum? Was würdest du dort machen und wen würdest du mitnehmen?

('Where would you most likely like to travel and why? What would you do there and who would you take along?')

c) Wer ist dein/e beste/er Freund/in und warum? Was findest du gut an deiner/em Freund/in?

('Who is your best boyfriend or girlfriend and why? What is it you like about your friend?')

Part III:

Hier ist ein Bilderbuch. Die Geschichte vom Frosch im Restaurant von Mercer Mayer. Bitte beschreibe was du siehst und erzähle mir die Geschichte.

('Here is a picture book. The story about the frog in a restaurant by Mercer Mayer. Please describe what you see and tell me the story.')

Part IV:

a) Wie findest du die Geschichte? Und welche Stelle hat dir am besten gefallen?

('What do you think about the story? And which part did you like the best?')

b) Kennst du Mercer Mayer und weißt du, was er noch geschrieben hat?

('Do you know Mercer Mayer and do you know what else he wrote?')

c) Wie fandest du das Deutschsprechen, die Fragen und alles zusammen? War es schwer oder einfach?

('What did you think about this speaking German, the questions, and everything? Was it difficult or easy?')

APPENDIX G

Rating and Letter Grading Scale for German-English Bilingual Grade Six Students

Adapted from the SOPA

	JR NOVICE LOW	JR NOVICE MID	JR NOVICE HIGH
Listening Comprehension	Recognizes isolated words and high frequency expressions.	Understands predictable questions, statements, and commands in familiar topic areas (with strong contextual support), though at slower than normal rate of speech and/or with repetitions.	Understands simple questions, statements and commands in familiar topic areas, and some new sentences with strong contextual support. May require repetition, slower speech, or rephrasing.
Fluency	Produces only isolated words in very specific topic areas and/or high-frequency expressions such as *good morning*. Has essentially no functional communicative ability.	Uses a limited number of words and phrases within predictable topic areas. May also use longer memorized expressions.	Uses high frequency expressions and other memorized expressions with reasonable ease. Signs of originality and spontaneity begin to emerge. Attempts to create sentences with some success, but is unable to sustain sentence level speech.
Vocabulary	Uses words in very specific topic areas in predictable contexts and a few memorized, high frequency expressions.	Uses vocabulary consisting of specific words, high-frequency expressions, and/or other longer, memorized expressions in a limited number of topic areas. Frequent searches for words are common.	Uses vocabulary centering on basic objects, places, and common kinship terms, adequate for minimally elaborating utterances in predictable topic areas.
Grammar	May use memorized, high-frequency phrases accurately, but lacks awareness of grammar and syntax.	May use memorized expressions accurately. Attempts at putting together two or three word phrases may also be accurate. Does not use verbs to speak creatively in sentence-level speech.	Relies mainly on memorized expressions, and may use them accurately. Creates some sentences-level speech with conjugated verbs, but in general, verbs are lacking or are not conjugated.

Note: Any score within the Novice Level is considered a failing Grade D (Work does not meet acceptable standard).

	JR INTERMEDIATE LOW	JR INTERMEDIATE MID	JR INTERMEDIATE HIGH
Listening Comprehension	Understands new sentence-level questions and commands in a limited number of content areas with strong contextual support. Follows conversation at a fairly normal rate of speech. C	Understands sentence-level utterances in new contexts though some slow-downs may be necessary. Carries out commands without prompting. May show some difficulty on unfamiliar topics. B-	Usually understands speech at normal speed. Seldom has comprehension problems on everyday topics. Understands longer stretches of connected discourse on a number of topics. B+
Fluency	Able to handle a limited number of everyday social and academic interactions. Maintains simple conversations at the sentence level by creating with the language, although in a restrictive and reactive manner. D	Shows evidence of spontaneity in conversation. May initiate talk without relying on questions or prompts. Gives simple descriptions successfully. May attempt longer, increasingly complex sentences. Few, if any connectors are used. C	Maintains conversation with increasing fluency. Uses language creatively to initiate and sustain talk. There is emerging evidence of connected discourse in descriptions and simple narratives. B
Vocabulary	Has sufficient vocabulary for making statements and asking questions to satisfy basic social and academic needs, but not for explaining or elaborating on them. D	Has basic vocabulary permitting discussion of a personal nature and limited academic topics. Serious gaps may exist for speaking about topics of general interest. Attempts at circumlocution may be ineffective. C	Has a broad enough vocabulary for discussing simple social and academic topics in generalities, but lacks detail. Sometimes achieves successful circumlocution. B
Grammar	Goes beyond memorized utterances to create with the language at the sentence level. Verbs are conjugated, but may be inaccurate. Speech may also contain many other grammatical inaccuracies. C-	Maintains simple conversation, mostly in present tense, although an awareness of other tenses may be evident. Many grammatical inaccuracies are present. C	Uses present tense, but lacks control of past tenses. Grammatical inaccuracies are present. B

	JR ADVANCED	JR ADVANCED-HIGH	SUPERIOR
Listening Compre-hension	Understands main ideas and most details in connected speech on a variety of topics, and is aware of connectors, but may be unable to follow complicated oral speech. (May have difficulty with highly idiomatic speech.)* **A**	Understands complex academic speech and highly idiomatic conversation, though confusion may occur due to sociocultural nuances or unfamiliar topics.	Has no difficulty understanding social or academic interactions. Follows extended speech in lectures, speeches, and reports, having an awareness of underlying organizational patterns.
Fluency	Reports facts easily. Can discuss topics of personal interest and some topics, satisfying the requirements of school and every day situations. Narrates and describes successfully. Connects sentences smoothly, and organizes speech using sequencing devices such as the first, next, finally, etc. **A**	Handles most academic and social requirements with confidence, but may demonstrate inadequacy under the demands of complex tasks. Emerging ability to support opinions and hypothesize on abstract topics is evident.	Uses a variety of interactive strategies to participate effectively in formal and informal, social and academic situations. Can support opinions, hypothesize on abstract concepts, and adjust speech to the type of audience.
Vocabulary	Has vocabulary adequate for including some detail when talking about concrete or factual topics of a personal nature and of general interest. (May use a variety of idiomatic expressions.)* Can use circumlocution effectively. **A-**	Has sufficient vocabulary to discuss a wide variety of topics related to everyday social and academic situations. Lack of vocabulary rarely interrupts the flow of speech. **A+**	Is familiar with idiomatic expressions and less common vocabulary which permit discussing topics in unfamiliar situations. Vocabulary is extensive.
Grammar	Has good control in expressing present, past, and future times. Errors, if present, do not interfere with communication. **A-**	Uses conditionals, but can not sustain a hypothetical mode with appropriate verb forms. Uses most grammatical structures accurately, but patterns of error may persists. **A+**	Uses complex forms as well. Occasional errors in complex, low-frequency structures may occur, but they do not interfere with communication. There are no major patterns of error.

* Criteria in parentheses do not apply for this age group.

Note: Unmarked scores within Jr. Advanced-High and Superior exceed expectations for Grade Six and result in a Grade A+.

APPENDIX H

Field Testing including all Four Sections

(All of the questions were used as guidelines. The actual wording could be different.)

Section I.: Warm-up

1. Bevor wir anfangen, sag mir doch bitte, wie du heißt, wie alt du bist, wie viele Geschwister du hast, wo du wohnst, und was dein Lieblingsessen ist.

   (Before we start, please tell me your name, how old you are, how many brothers and sisters you have, where you live, and what your favorite food is.)

Section II.: Short answers

A. Mainly present tense expected:

a) Repetitive tasks:

2. Kannst du mir bitte erklären, wie man das Folgende macht: Wie leiht man ein Buch aus der Schulbibliothek aus?

   (Can you please explain to me, how you do the following: how does one borrow a book from the school library?)

3. Was machst du jeden Tag, von dem Zeitpunkt ab, an dem du aufstehst, bis du abends ins Bett gehst?

   (What do you do every day, from the moment you get up, until you go to bed in the evening?)

4. Wie sieht ein Montag/Dienstag, usw. in der Schule aus?

   (What does a Monday/Tuesday, etc. look like at school?)

5. Was machst du meistens am Wochenende?

   (What do you mostly/usually do on the weekend?)

6. Machst/treibst du Sport? Wenn ja, was machst du und kannst du mir den Sport beschreiben? Wenn nein, hast du ein Hobby und kannst du es mir beschreiben?

   (Do you do any sports? If yes, what do you do and can you describe the sport to me? If no, do you have a hobby and can you describe it to me?)

b) Descriptions:

(Beginning/middle of school year)

Easy:

7. Kannst du dich beschreiben? Was du (nicht) gerne machst, was du (nicht) gerne isst, was du dir wünschst, wie du bist, usw.

   (Can you describe yourself? What you do/do not like, what you do/do not like to eat, what you wish for, how you are, etc.)

8. Kannst du mir bitte deine Familie beschreiben? Aussehen und was alle mögen, nicht mögen, usw.

   (Can you please describe your family to me? Appearance and what they all like, not like, etc.)

9. Kannst du mir bitte deine/n beste/n Freund/in beschreiben? Aussehen und was er/sie mag und was nicht.

   (Can you please describe your best friend to me? Appearance and what he/she likes and what not.)

10. Hast du ein Haustier? Kannst du es mir ganz genau beschreiben? Wie es aussieht, was es macht, was es gerne isst, warum du es magst?

    (Do you have a pet? Can you describe it to me well? What it looks like, what it does, what it likes to eat, why you like it?)

11. Kannst du mir bitte unsere Schule beschreiben? Wie sie aussieht, was es hier für Regeln gibt, was man hier alles macht? Was man nicht machen darf.

    (Can you please describe your school? What it looks like, what the rules are, all the things one can do here? What one is not allowed to do?)

12. Was machst du am liebsten in deiner Freizeit?

    (What do you like to do most in your free time?)

(Year end)

More difficult:

13. Kannst du mir bitte beschreiben, wie ein Fahrrad aussieht? Ein Auto? Deine Wohnung, oder dein Haus? Dein Zimmer?

    (Can you please describe, what a bicycle looks like? A car? Your apartment or house? Your room?)

14. Kannst du mir bitte dieses Bild beschreiben? Was siehst du? Was magst du und was magst du daran nicht?

    (Can you please describe this piece of art to me? What do you see? What do you like and what don't you like about it?)

15. Kannst du mir bitte Edmonton beschreiben? Was kann man hier alles machen und sehen?

    (Can you please describe Edmonton to me? What can one do and see here?)

16. Kannst du mir bitte beschreiben, wie man hier in Kanada folgende Feste feiert: Weihnachten, Ostern, Geburtstage.

    (Can you please describe to me how one celebrates the following holidays here in Canada: Christmas, Easter, Birthdays.)

17. Kannst du mir bitte China beschreiben? Die Leute, das Land, das Essen, usw.

    (Can you please describe China to me? The people, the country, the food, etc.)

18. Kannst du mir bitte Albertas Bodenschätze beschreiben? Welche gibt es und wo?

    (Can you please describe Alberta's natural resources to me? Which ones are there and where?)

Complex situations; concrete:

19. Stell dir vor, du bist in einem Restaurant in Deutschland und der Kellner/die Kellnerin hat dir das falsche Essen gebracht. Wie beschwerst du dich?

    (Imagine, you were in a restaurant in Germany and the waiter/waitress brought you the wrong food. How do you lodge a complaint?)

20. Du hast dich mit deiner/deinem besten Freundin/Freund gestritten und es ist deine Schuld. Wie entschuldigst du dich jetzt? Was sagst/machst du?

    (You had an argument with your best friend and it's your fault. How do you excuse yourself? What do you say/do?)

21. Jemand (ein Freund, ein Mitschüler) will dir Zigaretten/Alkohol/Drogen verkaufen. Wie verhältst du dich? Was sagst/machst du?

    (Someone, a friend, another student, wants to sell you cigarettes/alcohol/drugs. How do you react? What do you say/do?)

B. Mainly past tense expected (simple past; present perfect):

a) Descriptions:

(Beginning/middle of school year)

Easy to more difficult:

22. Was hast du früher mal gemacht, was du jetzt nicht mehr machst? Z. B.: eine Sportart; etwas Sammeln; eine Angewohnheit; Musik, die du gehört hast; Essen, das du jetzt nicht mehr magst, usw.

    (What did you do in the past that you don't do anymore? For example: sports, collecting something, a habit, music you used to listen to, food that you don't like anymore, etc.)

23. Kannst du mir beschreiben, wo du mal im Urlaub/in den Ferien warst? Wo? Was hast du dort gesehen, gemacht? Was hat dir (nicht) gefallen?

    (Can you describe to me, where you once spend a vacation? Where? What did you see/do there? What did/didn't you like?)

24. Kannst du mir erzählen, was dir mal Tolles/Schlechtes passiert ist? Irgendetwas von früher, letzter Woche, letztem Jahr oder so.

    (Can you tell me about something great/terrible that happened to you once? Something from the past, last week, last year or so)

(Year end)

More difficult:

25. Was weißt du über die alten Griechen? Wie haben sie gelebt? Was haben sie gemacht?

    (What do you know about the ancient Greeks? How did they live? What did they do?)

C. Mainly future tense expected (forms of '*werden*' + infinitive or present tense):

a) Descriptions:

(Beginning/middle of school year)

Easy:

26. Was wirst du am Wochenende machen?

    (What will you do this weekend?)

27. Was wirst du in den Schulferien machen? Weihnachten? Im Sommer?
    (What will you do in your holidays? Christmas? In the summer?)

28. Wohin wirst du als nächstes reisen? Was wirst du da machen? Wenn du noch nie da warst, wie stellst du es dir da vor? Was man da machen kann und was nicht?

    (Where will you travel to next? What will you do there? If you haven't been there yet, what do you imagine it like? What can one do there and what not?)

(Year end)

More difficult:

29. Was willst du machen, wenn du mit der Schule fertig bist? Was willst du in der Zukunft werden/machen/arbeiten?

    (What will you do when you're done with school? What do you want to become/do/work in the future?)

30. Wohin möchtest du unbedingt mal reisen und warum? Was willst du da (nicht) machen?

    (Where do you want to travel to badly and why? What do/don't you want to do there?)

31. Jemand kommt das erste Mal in deine Schule und kommt zur Tür bei Mrs. A Klassenzimmer rein. Wie erklärst du den Weg zu Mrs. B? Zum Klo? Zur Bibliothek? Zum Musikzimmer?

    (Somebody comes to your school for the first time and enters through the door close to Mrs. A's classroom. How do you explain the way to Mrs. B? The washroom? The Library? The music room?)

32. Jemand kommt das erste Mal in deine Schule und steht bei Mrs. Bs Büro. Wie erklärst du den Weg zum Hausmeister? Zum Klo? Zum Lehrerzimmer? Zu Frau Cs Klassenzimmer? Zu deinem Klassenzimmer?

    (Somebody comes to your school for the first time and stands in front of Mrs. B's office. How do you explain the way to the janitor? The washroom? The staff room? Frau C's classroom? Your classroom?)

33. Jemand kommt zu dir nach Hause und ist draußen im Garten. Wie erklärst du den Weg zum Klo? Zur Küche? Zu deinem Zimmer? In den Keller?

    (Somebody comes to your home and is outside in the yard. How do you explain the way to the washroom? The kitchen? Your room? The basement?)

D. All tenses possible:

Comparisons:

(Beginning/middle of school year)

Easy:

34. Kannst du vergleichen, wie man Weihnachten in Deutschland und Kanada feiert? Was ist anders/ähnlich/gleich?

    (Can you compare how Christmas is celebrated in Germany and Canada? What is different/similar/the same?)

35. Kannst du vergleichen, wie man Halloween in Kanada und Fasching in Deutschland feiert? Was ist anders/ähnlich/gleich?

    (Can you compare how Halloween and carnival are celebrated in Canada and Germany respectively? What is different/similar/the same?)

(Year end)

More difficult:

36. Kannst du mir über die alten Griechen erzählen? Kannst du ihr Leben zum Leben in Kanada vergleichen?

    (Can you tell me about the ancient Greeks? Can you describe their life to life in Canada?)

37. Wie wohnen die Chinesen? Und kannst du ihr Leben zum Leben in Kanada vergleichen?

    (How do the Chinese live? And can you describe their life to life in Canada?)

E. Hypothetical, „what if" situations:

Was wäre, wenn du :

(What if you)

38. eine Millionen Dollar gewinnen würdest? Was würdest du mit dem Geld machen?

    (won one million dollar? What would you do with the money?)

39. nach China/Deutschland ziehen müsstest? Was würdest du hier vorher noch machen? Und was würdest du dort (nicht) machen?

    (had to move to China/Germany? What would you still do here before? And what would/wouldn't you do there?)

40. von jemandem angesprochen werden würdest, der dir Zigaretten/Alkohol/Drogen verkaufen wollte? Was würdest du sagen/machen?

    (were approached by somebody who wanted to sell you cigarettes/alcohol/drugs? What would you say/do?)

41. deine/n Lieblingsrockstar/-gruppe/-schauspieler kennenlernen würdest? Was würdest du machen/sagen/denken?

    (met your favourite rock star/rock group/actor? What would you do/say/think?)

42. auf dem Schulhof von jemandem angesprochen wirst, der mit dir streiten/kämpfen wollte? Was würdest du machen/sagen?

    (were approached by somebody in the schoolyard who wanted to argue/fight with you? What would you do/say?)

43. zur Zeit der alten Griechen leben würdest? Wie würdest du leben, was würdest du (nicht) machen?

    (lived during the time of the ancient Greeks? How would you live, what would/wouldn't you do?)

Section III.: Picture book narration

Hier ist ein Bilderbuch von Mercer Mayer. Die Geschichte handelt von einem Frosch und seinen Abenteuern. Kannst du mir bitte die Geschichte erzählen und beschreiben, was hier passiert?

(Here is a picture book by Mercer Mayer. The story is about a frog and his adventures. Could you please narrate the story to me and describe what happens here?)

Section IV.: Cool-down

44. Wie fandest du denn die Geschichte? Hat sie dir gefallen?

    (What did you think about the story? Did you like it?)

45. Das war wirklich sehr gut und wir sind jetzt fertig. Wie findest du die Geschichte?

    (That was really very good and we're done now. What do you think about the story?)

46. Wir sind fertig, wie fühlst du dich jetzt? War es anstrengend?

    (We're done, how do you feel now? Was it hard?)

APPENDIX I

Grammar Checklist

Expected to be known and used:

Tenses:

Present Tense	most verbs; modals (*müssen, können, dürfen, mögen, möchten, wollen, sollen*); *sein/haben*
Simple Past	most verbs; modals; *sein/haben*
Present Perfect	most verbs; *sein/haben*
Future	present tense (less commonly used: *werden* + infinitive)

Word order:

- Verb position in main clauses (and 'Time-Manner-Place')
- Verb position in Subordinate clauses
- Subordinate clauses with conjunctions *dass, weil*
- Imperatives (verb in initial position; singular, plural)
- Subject, Object, Indirect Object (by year end)
- Relative clauses in nominative (by year end)

Other grammar areas:

- Noun forms: singular, plural
- Formal address '*Sie*' (formal 'you')
- Imperatives
- Prepositions: accusative, dative, two-way prepositions
- Adjectives: comparatives/superlatives
- Possessive pronouns: nominative
- Adjective endings

Not expected to be known or used accurately, or not at all:

- Present Perfect: modals
- Past Perfect
- Future Perfect
- Subjunctive
- Genitive
- Possessive pronouns: accusative and dative

APPENDIX J

Additional Evaluation Criteria

Cultural knowledge about Germany

Able to elaborate on main ideas/comparisons, very good cultural knowledge.	A
Able to elaborate on main ideas/comparisons, good cultural knowledge.	B
Able to bring main points across, some cultural knowledge.	C
Little, if any cultural knowledge.	C-
Cultural knowledge is not satisfactory, not present.	D

Mixing with English (does not include unfamiliar vocabulary)

Some nouns only. Strong attempt to speak only German.	A
Some nouns, verbs, adjectives, adverbs. Attempts to speak only German.	B
Many nouns, verbs, etc, plus clauses. Little attempt to speak only German.	C
Clauses and complete sentences. No attempt to speak only German.	D

APPENDIX K

Final Test

(All of the questions are guidelines. The actual wording can be different.)

Section I.: Warm-up

1. Bevor wir anfangen, sag mir doch bitte, wie du heißt, wie alt du bist, wie viele Geschwister du hast, wo du wohnst, und was dein Lieblingsessen ist.

   (Before we start, please tell me your name, how old you are, how many brothers and sisters you have, where you live, and what your favorite food is.)

Section II.: Short answers

A. Mainly present tense expected:

a) Repetitive tasks:

2. Kannst du mir bitte erklären, wie man das Folgende macht: Wie leiht man ein Buch aus der Schulbibliothek aus?

   (Can you please explain to me, how you do the following: how does one borrow a book from the school library?)

3. Was machst du jeden Tag, von dem Zeitpunkt ab, an dem du aufstehst, bis du abends ins Bett gehst?

   (What do you do every day, from the moment you get up, until you go to bed in the evening?)

4. Wie sieht ein Montag/Dienstag, usw. in der Schule aus?

   (What does a Monday/Tuesday, etc. look like at school?)

5. Was machst du meistens am Wochenende?

   (What do you mostly/usually do on the weekend?)

6. Machst/treibst du Sport? Wenn ja, was machst du und kannst du mir den Sport beschreiben? Wenn nein, hast du ein Hobby und kannst du es mir beschreiben?

   (Do you do any sports? If yes, what do you do and can you describe the sport to me? If no, do you have a hobby and can you describe it to me?)

b) Descriptions:

(Beginning/middle of school year)

Easy:

7. Kannst du dich beschreiben? Was du (nicht) gerne machst, was du (nicht) gerne isst, was du dir wünschst, wie du bist, usw.

   (Can you describe yourself? What you do/do not like, what you do/do not like to eat, what you wish for, how you are, etc.)

8. Kannst du mir bitte deine Familie beschreiben? Aussehen und was alle mögen, nicht mögen, usw.

   (Can you please describe your family to me? Appearance and what they all like, not like, etc.)

9. Kannst du mir bitte deine/n beste/n Freund/in beschreiben? Aussehen und was er/sie mag und was nicht.

   (Can you please describe your best friend to me? Appearance and what he/she likes and what not.)

10. Hast du ein Haustier? Kannst du es mir ganz genau beschreiben? Wie es aussieht, was es macht, was es gerne isst, warum du es magst?

    (Do you have a pet? Can you describe it to me well? What it looks like, what it does, what it likes to eat, why you like it?)

11. Kannst du mir bitte unsere Schule beschreiben? Wie sie aussieht, was es hier für Regeln gibt, was man hier alles macht? Was man nicht machen darf.

    (Can you please describe your school? What it looks like, what the rules are, all the things one can do here? What one is not allowed to do?)

12. Was machst du am liebsten in deiner Freizeit?

    (What do you like to do most in your free time?)

(Year end)

More difficult:

13. Kannst du mir bitte beschreiben, wie deine Wohnung/dein Haus/dein Zimmer aussieht?

    (Can you please describe, what your apartment/your house/your room looks like?)

14. Kannst du mir bitte dieses Bild beschreiben? Was siehst du? Was magst du und was magst du daran nicht?

    (Can you please describe this piece of art to me? What do you see? What do you like and what don't you like about it?)

15. Wenn du Besuch aus Deutschland bekommst, was könnt ihr alles in Edmonton machen? Wohin könnt ihr gehen? Was kannst du zeigen?

    (If you have visitors from Germany, what can you do together in Edmonton? Where can you go? What can you show?)

16. Kannst du mir bitte beschreiben, wie man hier in Kanada folgende Feste feiert: Weihnachten, Ostern, Geburtstage.

    (Can you please describe to me how one celebrates the following holidays here in Canada: Christmas, Easter, Birthdays)

17. Kannst du mir bitte China beschreiben? Die Leute, das Land, das Essen, usw.

    (Can you please describe China to me? The people, the country, the food, etc.)

18. Kannst du mir bitte etwas über Albertas Bodenschätze erzählen? Welche gibt es und wo?

    (Can you please tell me something about Alberta's natural resources? Which ones are there and where?)

Complex situations; concrete:

19. Du hast dich mit deiner/deinem besten Freundin/Freund gestritten und es ist deine Schuld. Wie entschuldigst du dich jetzt? Was sagst/machst du?

    (You had an argument with your best friend and it's your fault. How do you excuse yourself? What do you say/do?)

20. Jemand (ein Freund, ein Mitschüler) will dir Zigaretten/Alkohol/Drogen verkaufen. Wie verhältst du dich? Was sagst/machst du?

    (Someone, a friend, another student, wants to sell you cigarettes/alcohol/drugs. How do you react? What do you say/do?)

B. Mainly past tense expected (simple past; present perfect):

a) Descriptions:

Easy to more difficult:

21. Was hast du früher mal gemacht, was du jetzt nicht mehr machst? Z. B.: eine Sportart; etwas Sammeln; eine Angewohnheit; Musik, die du gehört hast; Essen, das du jetzt nicht mehr magst, usw.

    (What did you do in the past that you don't do anymore? For example: sports, collecting something, a habit, music you used to listen to, food that you don't like anymore, etc.)

22. Kannst du mir beschreiben, wo du mal im Urlaub/in den Ferien warst? Wo? Was hast du dort gesehen, gemacht? Was hat dir (nicht) gefallen?

    (Can you describe to me, where you once spend a vacation? Where? What did you see/do there? What did/didn't you like?)

23. Kannst du mir erzählen, was dir mal Tolles/Schlechtes passiert ist? Irgendetwas von früher, letzter Woche, letztem Jahr oder so.

    (Can you tell me about something great/terrible that happened to you once? Something from the past, last week, last year or so)

24. Was weißt du über die alten Griechen? Wie haben sie gelebt? Was haben sie gemacht?

    (What do you know about the ancient Greeks? How did they live? What did they do?)

C. Mainly future tense expected (present tense mostly; possibly forms of '*werden*' + infinitive):

a) Descriptions:

(Beginning/middle of school year)

Easy:

25. Was wirst du am Wochenende machen?

    (What will you do this weekend?)

26. Was wirst du in den Schulferien machen? Weihnachten? Im Sommer?

    (What will you do in your holidays? Christmas? In the summer?)

27. Wohin wirst du als nächstes reisen? Was wirst du da machen? Wenn du noch nie da warst, wie stellst du es dir da vor? Was man da machen kann und was nicht?

    (Where will you travel to next? What will you do there? If you haven't been there yet, what do you imagine it like? What can one do there and what not?)

(Year end)

More difficult:

28. Was willst du machen, wenn du mit der Schule fertig bist? Was willst du in der Zukunft werden/machen/arbeiten?

    (What will you do when you're done with school? What do you want to become/do/work in the future?)

29. Wohin möchtest du unbedingt mal reisen und warum? Was willst du da (nicht) machen?

    (Where do you want to travel to badly and why? What do/don't you want to do there?)

30. Jemand kommt das erste Mal in deine Schule und kommt zur Tür bei Mrs./Frau As Klassenzimmer rein. Wie erklärst du den Weg zu Mrs./Frau B? Zum Klo? Zur Bibliothek? Zum Musikzimmer?

    (Somebody comes to your school for the first time and enters through the door close to Mrs./Frau A's classroom. How do you explain the way to Mrs./Frau B? The washroom? The library? The music room?)

31. Jemand kommt das erste Mal in deine Schule und steht bei Mrs. Cs Büro. Wie erklärst du den Weg zum Hausmeister? Zum Klo? Zum Lehrerzimmer? Zu Mrs./Frau Ds Klassenzimmer? Zu deinem Klassenzimmer?

    (Somebody comes to your school for the first time and stands in front of Mrs. C's office. How do you explain the way to the janitor? The washroom? The staff room? Mrs./Frau D's classroom? Your classroom?)

32. Jemand kommt zu dir nach Hause und ist draußen im Garten. Wie erklärst du den Weg zum Klo? Zur Küche? Zu deinem Zimmer? In den Keller?

    (Somebody comes to your home and is outside in the yard. How do you explain the way to the washroom? The kitchen? Your room? The basement?)

D. All tenses possible:

Comparisons:

(Beginning/middle of school year)

Easy (German culture):

33. Kannst du vergleichen, wie man Weihnachten in Deutschland und Kanada feiert? Was ist anders/ähnlich/gleich?

    (Can you compare how Christmas is celebrated in Germany and Canada? What is different/similar/the same?)

34. Kannst du vergleichen, wie man Halloween in Kanada und Fasching in Deutschland feiert? Was ist anders/ähnlich/gleich?

    (Can you compare how Halloween and carnival are celebrated in Canada and Germany respectively? What is different/similar/the same?)

35. Kannst du erklären, wie man, im Vergleich zu Deutschland in Kanada wohnt? Wie sind die Häuser anders?

    (Can you explain how one lives in Canada, in comparison to Germany? In what way are the houses different?)

(Year end)

More difficult:

36. Kannst du mir über die alten Griechen erzählen? Kannst du ihr Leben zum Leben in Kanada vergleichen?

    (Can you tell me about the ancient Greeks? Can you describe their life to life in Canada?)

37. Wie wohnen die Chinesen? Und kannst du ihr Leben zum Leben in Kanada vergleichen?

    (How do the Chinese live? And can you describe their life to life in Canada?)

38. War das Leben in China vor 1979 besser oder schlechter als jetzt?

    (Was life in China in 1979 better or worse than today?)

E. Hypothetical, "what if" situations:

Was wäre, wenn du:

(What if you)

39. eine Millionen Dollar gewinnen würdest? Was würdest du mit dem Geld machen?

    (won one million dollars? What would you do with the money?)

40. nach China/Deutschland ziehen müsstest? Was würdest du hier vorher noch machen? Und was würdest du dort (nicht) machen?

    (had to move to China/Germany? What would you still do here before? And what would/wouldn't you do there?)

41. von jemandem angesprochen werden würdest, der dir Zigaretten/Alkohol/Drogen verkaufen wollte? Was würdest du sagen/machen?

    (were approached by somebody who wanted to sell you cigarettes/alcohol/drugs? What would you say/do?)

42. deine/n Lieblingsrockstar/-gruppe/-schauspieler/-sportler kennenlernen würdest? Was würdest du machen/sagen/denken?

    (met your favourite rock star/rock group/actor/sports celebrity? What would you do/say/think?)

43. auf dem Schulhof von jemandem angesprochen wirst, der mit dir streiten/kämpfen wollte? Was würdest du machen/sagen?

    (were approached by somebody in the schoolyard who wanted to argue/fight with you? What would you do/say?)

Section III.: Picture description

Kannst du mir bitte beschreiben, was du auf diesem Bild siehst?

(Could you please describe what you see in this picture?)

Section IV.: Cool-down

44. Wir sind fertig, wie fühlst du dich jetzt? War es anstrengend?

    (We're done, how do you feel now? Was it hard?)